Архив 2016-2023 гг.

Осенний семестр 2023 г.

Дата	Докладчик и тема	Материалы
21.11.2023	Timur Garipov (PhD student, Computer Science at MIT EECS & MIT CSAIL) Compositional Sculpting of Iterative Generative Processes High training costs of generative models and the need to fine-tune them for specific tasks have created a strong interest in model reuse and composition. A key challenge in composing iterative generative processes, such as GFlowNets and diffusion models, is that to realize the desired target distribution, all steps of the generative process need to be coordinated, and satisfy delicate balance conditions. In this work, we propose Compositional Sculpting: a general approach for defining compositions of iterative generative processes. We then introduce a method for sampling from these compositions built on classifier guidance. We showcase ways to accomplish compositional sculpting in both GFlowNets and diffusion models. We highlight two binary operations — the harmonic mean (p ⊗ q) and the contrast (p ◑ q) between pairs, and the generalization of these operations to multiple component distributions. We offer empirical results on image and molecular generation tasks.	Презентация **Статьи:** Compositional Sculpting of Iterative Generative Processes Compositional Visual Generation and Inference with Energy Based Models Reduce, Reuse, Recycle: Compositional Generation with Energy-Based Diffusion Models and MCMC Multi-Objective GFlowNets

Весенний семестр 2023 г.

Дата	Докладчик и тема	Материалы
10.02.2023	Екатерина Рахилина (НИУ ВШЭ) Все дело в языке, а он – меняется. И поэтому... В докладе речь пойдет о языковых изменениях. Это постоянный и неуправляемый процесс, связанный с изменением значений большого числа переменных. Однако, прежде чем научиться его измерять, вычислять и предсказывать, нужно убедиться в том, что он есть и масштабен – причем не только на длительных исторических промежутках (латынь и современный французский), но и на более обозримых для сегодняшних говорящих. Как правило, в это носителям языка трудно поверить – но мы в этом убедимся на интересных примерах.	**Видео**
03.03.2023	Bayram Akdeniz (Norwegian Centre for Mental Disorders Research), Oleksandr Frei (University of Oslo)	Презентация 1 (Байрам Акдениз) Презентация 2 (Байрам Акдениз & Олександр Фрей)

	Finemap-MiXeR: A variational Bayesian approach for genetic finemapping Genome-wide association studies (GWAS) implicate large clusters of highly correlated genetic variants, which makes it hard to interpret the results from a biological point of view. Several methods in statistical genetics allow to "finemap" underlying causal variants, or at least point to genes or gene sets that are responsible for associations observed in GWAS. In the seminar we will discuss two new methods, GSA-MiXeR and Finemap-MiXeR, which address these issues. Finemap-MiXeR is a variational Bayesian approach for finemapping genomic data, using optimization of Evidence Lower Bound of the likelihood function obtained from the MiXeR model. The optimization is done using Adaptive Moment Estimation Algorithm, allowing to obtain posterior probability of each SNP to be a causal variant. We apply Finemap-MiXeR to a range of different scenarios, using both synthetic and real data from the UK Biobank, using standing height phenotype as an example. In comparison to the existing finemapping methods FINEMAP and SuSiE methods, we observed that Finemap-MiXeR in most cases has better accuracy. For the GSA-MiXeR we'll demonstrate its application to schizophrenia, where GSA-MiXeR implicate the role of calcium channel function, GABAergic and dopaminergic signaling. To conclude, we will discuss more broadly what are some of the main methodological challenges in statistical genetics, and how new methods can improve our understanding of complex polygenic traits and disorders.	**Статьи**: Finemap-MiXeR: A variational Bayesian approach for genetic finemapping Improved functional mapping with GSA-MiXeR implicates biologically specific gene-sets and estimates enrichment magnitude **Видео**
10.03.2023	Антон Вахрушев (Sber AI Lab) SketchBoost: быстрый бустинг для multiclass/multilabel классификации и multitask регрессии Градиентный бустинг – один из самых эффективных инструментов для решения задач машинного обучения на табличных данных. Однако в задачах, когда требуется прогнозировать сразу несколько выходов, таких как multiclass/multilabel классификация и multitask регрессия, построение бустинга на деревьях требует существенных вычислительных затрат и может занимать неприемлемо много времени. Мы придумали практичный метод сжатия информации, который применяется на каждой итерации бустинга, а также реализовали его на базе нашей библиотеки py-boost, которая доступна в opensource. В ходе нашего доклада мы расскажем, как можно добиться значительного ускорения времени обучения модели (в десятки раз) без каких-либо потерь в качестве.	Презентация **Статья:** SketchBoost: Fast Gradient Boosted Decision Tree for Multioutput Problems **Видео**
17.03.2023	Александр Панченко (Associate Professor, Skoltech, NLP Lab, AI Center) Monolingual and Cross-lingual Text Detoxification В этом докладе мы рассмотрим задачу переноса текстового стиля на примере задачи детоксикации текста. В первой части доклада мы рассмотрим моноязычный эксперимент сбора параллельных данных для задачи детоксикации. Мы собираем нетоксичные парафразы для английских и русских токсичных предложений. Используя полученный набор данных, мы обучаем несколько моделей seq2seq детоксикации на собранных данных и сравниваем их с несколькими базовыми моделями и современными подходами, не требующими наблюдения. Все модели, обученные на параллельных данных, с большим отрывом превосходят современные модели. Во второй части доклада мы рассмотрим многоязычный эксперимент, в котором мы решаем проблему детоксикации текста для языка, на котором отсутствует параллельный корпус. Кроме этого, мы обсудим эксперименты, в которых перевод и передача стиля должны решаться совместно.	Презентация Чат-бот с демо моноязычных (английский и русский) текстовых детоксификаторов Веб-приложения **Статьи:** ParaDetox: Detoxification with Parallel Data RUSSE-2022 Detoxification **Видео**
24.03.2023	Андрей Охотин (НИУ ВШЭ) Star-Shaped Denoising Diffusion Probabilistic Models Methods based on Denoising Diffusion Probabilistic Models (DDPM) became a ubiquitous tool in generative modelling. However, they are mostly limited to Gaussian and discrete diffusion processes. We propose Star-Shaped Denoising Diffusion Probabilistic	Презентация **Статья:** Star-Shaped Denoising Diffusion Probabilistic Models

	Models (SS-DDPM), a model with a non-Markovian diffusion-like noising process. In the case of Gaussian distributions, this model is equivalent to Markovian DDPMs. However, it can be defined and applied with arbitrary noising distributions, and admits efficient training and sampling algorithms for a wide range of distributions that lie in the exponential family. We provide a simple recipe for designing diffusion-like models with distributions like Beta, von Mises--Fisher, Dirichlet, Wishart and others, which can be especially useful when data lies on a constrained manifold such as the unit sphere, the space of positive semi-definite matrices, the probabilistic simplex, etc. We evaluate the model in different settings and find it competitive even on image data, where Beta SS-DDPM achieves results comparable to a Gaussian DDPM.	**Видео**
31.03.2023	Ильдус Садртдинов (НИУ ВШЭ) Self-supervised Pre-training with Masked Image Modeling Consistent success of BERT-like models in language processing has lead to attempts to adapt masked modeling task to other data domains and create a universal framework for pre-training without manual annotations. In this seminar, we will have an overview of relatively recent (2021-2022) approaches for masked image modeling (MIM). We will start with a brief history of self-supervised methods for images and then discuss different masking strategies, image-processing pipelines and what targets are suitable for masked modeling. We will consider BEiT, SimMIM, MAE, UM-MAE and MaskFeat.	Презентация **Статьи:** BEiT: BERT Pre-Training of Image Transformers SimMIM: A Simple Framework for Masked Image Modeling Masked Autoencoders Are Scalable Vision Learners Uniform Masking: Enabling MAE Pre-training for Pyramid-based Vision Transformers with Locality Masked Feature Prediction for Self-Supervised Visual Pre-Training **Видео**
07.04.2023	Айбек Аланов (AIRI, НИУ ВШЭ) Image Manipulation by Diffusion Models Large-scale text-to-image diffusion models have shown their impressive ability to synthesize diverse and high-quality images. However, it is still challenging to directly apply these models for editing real images or generating special concepts provided by the user. Recently there were proposed many works to deal with these problems. In this seminar, we will examine most of these approaches, analyze its properties and indicate their advantages and weak sides. At first, we will start with a brief overview of contemporary text-to-image diffusion models. Secondly, we will give the formulation for two types of image manipulation problems: personalized generation of the user-provided concept and editing of the given real image. Then we will examine current methods that tackle these problems.	Презентация **Статьи**: An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation Prompt-to-Prompt Image Editing with Cross Attention Control Imagic: Text-Based Real Image Editing with Diffusion Models Null-text Inversion for Editing Real Images using Guided Diffusion Models **Видео**
14.04.2023	Илья Трофимов (Skoltech) Topological data analysis and machine learning В последнее время были разработаны эффективные методы вычислительной топологии. Концепции, использовавшиеся ранее только в математике (например, персистентные гомологии), нашли применение в машинном обучении. Топология описывает "форму" данных на разных масштабах. Топологический "взгляд" дополняет существующие подходы к анализу данных. В данном докладе будет выполнен обзор нескольких результатов, полученных на стыке топологии и машинного	**Статьи:** Manifold Topology Divergence: a Framework for Comparing Data Manifolds Representation Topology Divergence: A Method for Comparing Neural Network Representations Learning Topology-Preserving Data Representations **Видео**

	обучения: для оценки генеративных моделей, сравнения эмбедингов нейросетей, топологически регуляризованного понижения размерности.	
21.04.2023	Никита Морозов (НИУ ВШЭ), Вячеслав Мещанинов (НИУ ВШЭ), Григорий Бартош (AMLab, UvA) Дискуссионный семинар о диффузных моделях Как и в предыдущем осеннем сезоне, в весеннем у нас будет один семинар в необычном формате. Выступят три докладчика с небольшими (по 20-25 минут) докладами на общую тему диффузных моделей. Мы поговорим о применении диффузных моделей для генерации 3d объектов. Обсудим техники, позволяющие добиться SOTA качество генерации изображений в высоком разрешении, без латентной диффузии. И разберем новый simulation-free подход к обучению ODE, вдохновленный диффузными моделями.	Презентация 1 (Никита Морозов) Презентация 2 (Вячеслав Мещанинов) Презентация 3 (Григорий Бартош) **Статьи:** Novel View Synthesis with Diffusion Models Simple diffusion: End-to-end diffusion for high resolution images Flow Matching for Generative Modeling **Видео**
28.04.2023	Нет семинара	
12.05.2023	Михаил Самин (CEO, AudD) AI Alignment problem Не кажется непредставимым, что в течение следующего десятилетия, искусственный интеллект может превзойти способности людей в большинстве важных областей. Среди исследователей, работающих над созданием общего ИИ, есть растущее ожидание, что без усилий для предотвращения этого, ИИ могут выучить и преследовать цели, нежеланные с точки зрения людей, что может привести к катастрофическим последствиям. Десятки сотрудников OpenAI, Anthropic и DeepMind, работающих над проблемой, обычно называли мне числа от 15 до 80 процентов вероятности, что человечество будет буквально уничтожено искусственным интеллектом в ближайшие два десятилетия. Я постараюсь описать базовые причины, почему проблема алайнмента — как сделать так, чтобы цели достаточно продвинутого ИИ было приемлемыми — сложна, как выглядят некоторые независимые риски и какие есть направления решений.	Manifold **Видео**
02.06.2023	Александра Волохова (PhD Student, Mila — Quebec AI institute, Université de Montréal) Generative flow networks Generative flow networks (GFlowNets) are amortized variational inference algorithms that are trained to sample from unnormalized reward distributions over compositional objects. The key feature of these algorithms is their ability to generate a diverse set of high-reward objects, which is very useful for scientific discovery applications. Also, GFlowNets can accommodate both discrete and continuous properties of the objects. In this talk, I'll focus on explaining the algorithm and its mathematical foundations, its relation to diffusion models, and show you some examples of its applications.	**Статьи**: GFlowNet Foundations GFlowNets and variational inference Unifying Generative Models with GFlowNets and Beyond A theory of continuous generative flow networks Generative Flow Networks for Discrete Probabilistic Modeling Полный список статей в блог-посте Йошуа

Осенний семестр 2022 г.

Дата	Докладчик и тема	Материалы
16.09.2022	Александр Панов, Даниил Кириленко, Алексей Ковалев (AIRI) Discrete Disentangled Representations for Object-Centric Visual Tasks Recently, the pre-quantizing image features into discrete latent variables has helped achieve remarkable results in image modeling tasks. In this talk, we propose a method for using learnable discrete latent variables applied to object-centric tasks. Our approach utilizes the idea of slot object representation and models non-overlapping sets of low-dimensional discrete variables, sampling one vector from each to obtain the latent representation of the object. We empirically demonstrate that embeddings from the learned discrete latent spaces have the disentanglement property. The model exploits a set prediction as a downstream task and achieves the state-of-the-art results on the CLEVR dataset. We also apply it to the object discovery task and demonstrate manipulation of individual objects in the scene with controllable image generation. Дискретные распутанные представления для объектно-ориентированных визуальных задач В последнее время удалось достичь значительных результатов в задачах моделирования изображений используя предварительное квантование признаков изображения. В докладе предлагается метод, использующий обучаемые латентные дискретные представления для решения объектно-ориентированных задач. Предлагаемый подход развивает идею слотового представления объектов, моделируя непересекающиеся множества низкоразмерных дискретных представлений. Выбирая один вектор из каждого множества формируется скрытое представление объекта. Проведенные эксперименты демонстрируют, что выученные представления в дискретных скрытых пространствах являются распутанными. Такой подход был применён к задаче предсказания множества и показал лучший результат на наборе данных CLEVR по сравнению с неспециализированными моделями. Так же, применение этого подхода к задаче обнаружения объектов показало возможность управляемой генерации изображений сцены за счет манипулирования скрытыми представлениями отдельных объектов.	Презентация Видео
23.09.2022	Никита Стародубцев (ИТМО), Вячеслав Мещанинов (НИУ ВШЭ), Никита Бондарцев, Григорий Бартош (PhD Student at AMLab, UvA) Дискуссионный семинар о диффузных моделях На семинарах нашей группы мы уже не раз обсуждали диффузные модели - относительно новый класс генеративных моделей, активно набирающий популярность. В эту пятницу мы вновь вернемся к этой теме. Однако в этот раз семинар пройдет в дискуссионном формате. Будет четыре докладчика, каждый из которых расскажет про свою статью. Мы поговорим про генерацию изображений по текстовым описаниям, ускорение генерации с помощью ГАНов и диффузию на основе блюринга (размытия), а не стандартного зашумления. Так как формат дискуссионный, будет здорово, если перед семинаром вы пробежитесь по статьям и приготовите какие-то вопросы!	**Презентации:** Starodubtsev_GLIDE Bondartsev_Cold_Diffusion Meshchaninov__Tackling_the_Generative_Learning_Trilemma_with_Denoising_Diffusion **Статьи:** Editing with Text-Guided Diffusion Models Tackling the Generative Learning Trilemma with Denoising Diffusion GANs Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise

		Soft Diffusion: Score Matching for General Corruptions Blurring Diffusion Models **Видео**
30.09.2022	Нет семинара	
07.10.2022	Григорий Бартош (UvA), Аким Котельников (Яндекс, НИУ ВШЭ) Марковский блюрринг + диффузия для табличных данных Это продолжение дискуссионного семинара, посвященного диффузным моделям, который состоялся 23го сентября. Он будет состоять из двух частей. Мы закончим обсуждать блюрринг в диффузных моделях и немного обобщим эти подходы. Поговорим о том, как вообще из линейных моделей разрушения информации (зашумление, блюрринг, даунскейлинг и пр.) делать марковские процессы. Как с помощью этих моделей оценивать плотность в точке. Также коллеги из Яндекса расскажут о своей недавно вышедшей работе про диффузию для табличных данных.	**Презентации:** Bartosh_Blurring Bartosh_Linear_Diffusion_Models **Статьи:** Generative Modelling With Inverse Heat Dissipation paper: https://arxiv.org/abs/2206.13397 Soft Diffusion: Score Matching for General Corruptions paper: https://arxiv.org/abs/2209.05442 Blurring Diffusion Models paper: https://arxiv.org/abs/2209.05557 TabDDPM: Modelling Tabular Data with Diffusion Models paper: https://arxiv.org/abs/2209.15421 **Видео**
14.10.2022	Сергей Трошин (НИУ ВШЭ) Overview of control techniques for text and image models За последнее время появилось большое число предобученных генеративных моделей с текстовым управлением (GPT-3, Dall-E, …), на обучение которых потратили огромные ресурсы. Хочется понимать, как можно эффективно использовать данные модели, управлять процессом генерации в условиях ограниченности ресурсов на дообучение. Доклад будет посвящен обзору методов управления генеративными моделями. В первой части доклада мы рассмотрим методы эффективного дообучения предобученных моделей для текстов, обсудим методы in-context обучения, energy-based управления генерацией текстов. Во второй части доклада мы рассмотрим способы управления для text-to-image моделей, поговорим про редактирование изображений, персонализированную генерацию.	**Презентация** **Статьи:** https://arxiv.org/abs/2208.01066 https://arxiv.org/abs/2205.05638 https://arxiv.org/abs/2202.11705 https://arxiv.org/abs/2208.01626 https://dreambooth.github.io/ **Видео**

21.10.2022	Максим Кодрян (НИУ ВШЭ) Training Scale-Invariant Neural Networks on the Sphere Can Happen in Three Regimes A fundamental property of deep learning normalization techniques, such as batch normalization, is making the pre-normalization parameters scale invariant. The intrinsic domain of such parameters is the unit sphere, and therefore their gradient optimization dynamics can be represented via spherical optimization with varying effective learning rate (ELR), which was studied previously. However, the varying ELR may obscure certain characteristics of the intrinsic loss landscape structure. In this talk, we investigate the properties of training scale-invariant neural networks directly on the sphere using a fixed ELR. We discover three regimes of such training depending on the ELR value: convergence, chaotic equilibrium, and divergence. We study these regimes in detail both on a theoretical examination of a toy example and on a thorough empirical analysis of real scale-invariant deep learning models. Each regime has unique features and reflects specific properties of the intrinsic loss landscape, some of which have strong parallels with previous research on both regular and scale-invariant neural networks training. Finally, we demonstrate how the discovered regimes are reflected in conventional training of normalized networks and how they can be leveraged to achieve better optima.	Презентация Статья Видео
28.10.2022	Нет семинара	
04.11.2022	Станислав Дробышевский Вопросно-ответная сессия	Видео
11.11.2022	Илья Зиганшин (Физический факультет МГУ им. М.В. Ломоносова) Квантовая механика Доклад полупопулярно излагает основы квантовой механики. В рассказе разберем ее аксиоматику. Разберем формализм чистых и смешанных состояний, разрешим парадокс кота Шредингера на основе явления декогеренции и рассмотрим, как происходит переход между квантовой и классической статистикой. Будет изложена современная систематика интерпретаций квантовой механики. Выведем простейший вид неравенств Белла и рассмотрим за что дали Нобелевскую премию. Рассмотрим задачу о квантовой бомбе и эксперименты с отложенным выбором. Также будут ответы на вопросы слушателей.	**Презентация** Полезные ссылки: -почти вся современная теорфизика в полупопулярном виде https://disk.yandex.ru/i/2K5xRF_jftAPyA -самый современный курс по квантовой механике на русском языке(в рамках его разрешено множество парадоксов): https://teach-in.ru/course/density-matrix https://teach-in.ru/course/density-matrix-part2 Видео
18.11.2022	Тимофей Южаков (НИУ ВШЭ) Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets В прошлом году исследователи из OpenAI продемонстрировали интересный феномен, который они назвали “grokking”. Заключается он в следующем: нейросеть резко переходит от качества случайного угадывания к идеальному качеству, причём случается это сильно после точки оверфиттинга. Практики же обычно останавливают обучение сетей при первом намеке на переобучение. В чем же причина такого эффекта, и сколько на самом деле нужно учить сети? На семинаре обсудим результы ряда исследовательских групп, частично проливающих свет на данные вопросы, а также поделимся нашими экспериментами и результатами.	Презентация Статья

25.11.2022	Александр Новиков (DeepMind) Discovering faster matrix multiplication algorithms with reinforcement learning В данной работе мы применили модифицированную программу AlphaZero для поиска быстрых алгоритмов умножения матриц в символьном виде. Мы фокусируемся на поиске быстрых алгоритмов умножения матриц небольшого размера, например, 2x2, а затем используем найденные алгоритмы рекурсивно. В результате, в работе получилось уменьшить число скалярных умножений, которое требуется для умножения матриц разных размеров. Умножения матриц – это билинейная операция, и (как любую линейную операцию можно представить при помощи матрицы) ее можно представить при помощи трехмерного тензора. Низкоранговые разложения данного тензора соответствуют алгоритмам умножения матриц, а ранг разложения соответствует числу скалярных умножений. Таким образом, задача генерации алгоритмов умножения матриц трансформируется в эквивалентную задачу поиска низкоранговых разложений фиксированного тензора. Мы обучили AlphaZero искать эти разложения, применив такие приемы, как генерация синтетических данных, эксплуатация симметрий задачи, обучение одного агента раскладывать несколько разных тензоров одновременно, и использовать нейросетевую архитектуру, заточенную под особенности задачи. Результаты работы опубликованы в: Fawzi, Alhussein, et al. "Discovering faster matrix multiplication algorithms with reinforcement learning." Nature 610.7930 (2022): 47-53.	Презентация Статья Видео
02.12.2022	Тингир Бадмаев (НИУ ВШЭ) Диффузионные модели в латентном пространстве За последние пару лет диффузионные модели продемонстрировали отличное качество генерации, сопоставимое с sota-алгоритмами на основе GAN и VAE. Однако зачастую процесс диффузии происходит в пространстве данных, которые имеют большую размерность. Значит, вдобавок к итеративной схеме, мы получаем очень долгий процесс генерации. На спецсеминаре мы подробно разберем модель на основе VAE и cont. DDPM. Latent Score Generative Model использует процесс диффузии в латентном пространстве VAE, что упрощает обучение для DDPM, поскольку теперь DDPM необходимо моделировать не сложное распределение многомерных данных, а распределение близкое к праеру в VAE. Второй плюс - исходное пространство может иметь дискретную структуру. Score Based GM на основе диффузии с гауссовскими распределениями не умеют моделировать дискретные данные. Вторым примером использования диффузии в латентных переменных, мы рассмотрим Stable Diffusion.	Презентация Статья 1, 2 Видео
09.12.2022	Айбек Аланов (AIRI, НИУ ВШЭ) Domain Adaptation of GANs Современные модели ГАНов требуют больших датасетов высокого качества для успешного обучения, что является серьезным ограничением на практике. Мы рассмотрим методы доменной адаптации, которые позволяют обучить ГАН на доменах, которые представлены небольшим числом примеров. Основной подход	Презентация Статьи: Training Generative Adversarial Networks with Limited Data StyleGAN-NADA: CLIP-Guided Domain

	в этой задаче - это файнтьюнинг модели, обученной на большой выборке, на новый домен. В качестве такой модели мы будем рассматривать sota-модель StyleGAN, обученный на датасете лиц FFHQ. В докладе будет сделан обзор существующих методов доменной адаптации StyleGAN. Далее будут представлены наши результаты по тому, как можно уменьшить на порядки число дообучаемых параметров при файнтьюнинге StyleGAN и как это позволяет решать задачу мульти-доменной адаптации, когда мы хотим дообучить модель сразу на несколько доменов. Во второй части доклада будет предложен подробный анализ важности каждой компоненты архитектуры StyleGAN для доменной адаптации в зависимости от схожести целевого домена с исходным. Далее мы рассмотрим, как этот анализ позволяет улучшить существующие методы адаптации и открывает новые интересные свойства этих методов.	Adaptation of Image Generators HyperDomainNet: Universal Domain Adaptation for Generative Adversarial Networks StyleAlign: Analysis and Applications of Aligned StyleGAN Models Видео
16.12.2022	Максим Рябинин (Яндекс, НИУ ВШЭ) От GPT-3 до ChatGPT: обучение языковых моделей на инструкциях и человеческих оценках В 2020 году исследователи из OpenAI обнаружили, что большие языковые модели можно не дообучать на целевой задаче: достаточно подать в качестве контекста несколько примеров для этой задачи с ответами на них, а иногда хватает и вовсе текстовой инструкции. После этого научное сообщество стало активно развивать способы, позволяющие повысить качество работы языковых моделей в такой постановке, получившей название zero-shot/in-context learning. Недавний релиз ChatGPT показал, что адаптированные к in-context learning и выполнению инструкций языковые модели имеют большое количество потенциальных приложений, в том числе таких, для которых сбор обучающей выборки ранее считался необходимым. На семинаре мы обсудим ряд работ, излагающих ключевые подходы и направления исследований для улучшения работы языковых моделей в постановке in-context learning. Одним из таких направлений является устоявшаяся парадигма instruction finetuning: обучаясь на разнообразных наборах из формулировок задач, входных данных и ответов, языковые модели лучше следуют инструкциям даже для новых задач. Не обойдём вниманием и идею обучения с подкреплением на оценках текстов людьми, лежащую в основе ChatGPT и предшествовавшей ей InstructGPT.	Презентация Статьи: Finetuned Language Models Are Zero-Shot Learners Multitask Prompted Training Enables Zero-Shot Task Generalization Training language models to follow instructions with human feedback Improving alignment of dialogue agents via targeted human judgements Scaling Instruction-Finetuned Language Models Видео

Весенний семестр 2022 г.

Дата	Докладчик и тема	Материалы

25.02.2022	Александр Лобашев, Сколтех Formalism of quantum mechanics from the point of view of machine learning This talk will be devoted to an introduction to quantum mechanics from the point of view of machine learning. We briefly review the main ideas of classical mechanics and related concepts from Hamiltonian Monte Carlo methods: the Hamilton's equations and the principle of least action. After that, we introduce stochasticity into the Hamilton equations and, using the Fokker-Planck equation for the probability distribution function, we arrive at the Schrödinger equation. After introducing the basic axioms of quantum mechanics, we will discuss its main features, such as the noncommutativity of observables, the uncertainty principle, and entanglement. Finally, we will review the path integral formulation of quantum mechanics and see some analogies between the approximation of path integrals and ensemble methods in deep learning.	Презентация Modern Quantum Mechanics (первая глава) Видео
04.03.2022	Сергей Шумский, Руководитель лаборатории когнитивных архитектур МФТИ Свободная энергия и интеллект Доклад посвящен изложению принципа свободной энергии Фристона, как математической основы теории машинного обучения и машинного интеллекта. Из первых принципов выводятся основные постулаты теории обучения с подкреплением, являющейся базовой моделью сильного ИИ. Рассмотрены варианты model-free и model-based reinforcement learning.	Презентация Видео
11.03.2022	Diego Granziol, Huawei London AI Theory A Random Matrix Theory approach to Deep Learning optimisation and generalisation In this talk we consider a random matrix theory model for the mini-batch perturbation of the Hessian of the loss, which leads to analytical scaling rules (linear and square root) for SGD and Adam as a function of batch size respectively, which we show experimentally. We also present an analytical theory for the generalisation of Stochastic Weight Averaging under the framework of a Gaussian process model for gradient perturbations. Our theory shows the necessity of appropriate regularisation and large learning rates, noted in practice.	Презентация Видео
18.03.2022	Тингир Бадмаев, МГУ им. М.В. Ломоносова Denoising Diffusion Restoration Models Многие задачи по восстановлению изображений можно представить как линейные обратные задачи. Недавнее семейство подходов к решению этих проблем использует стохастические алгоритмы, которые семплируют из апостериорного распределения. Однако хорошие решения часто требуют обучения с учителем, ориентированного на конкретную задачу, для моделирования апостериорного распределения, в то время как unsupervised методы обычно полагаются на неэффективные итерационные методы. Мы разберем работу Denoising Diffusion Restoration Models, основанную на вариационном выводе и диффузионных моделях. Мы убедимся в универсальности DDRM для любой обратной линейной задачи, разберем эффективное построение линейных операторов для задачи повышения разрешения, колоризации, маскирования.	Презентация Статья 1, 2 https://github.com/bmml_sem/2022/Badmaev_DDRM_DDPM_NCSN.pdf Видео
25.03.2022	Илья Синильщиков и Евгений Бобров Mathematics of Multi-Antenna Transmission in 5G networks	Презентация Литература:

	С каждым новым стандартом на дальнейшее развитие сетей сотовой связи всё сильнее влияют используемые в системе математические алгоритмы. Оказывается, что построение эффективной базовой станции пятого поколения не обходится без решения широкого спектра задач прикладной математики, таких как невыпуклая и комбинаторная оптимизация, статистическое оценивание и даже машинное обучение. Причём на некоторые из них специфика области накладывает крайне жёсткие ограничения на время работы, что требует , например, разработки одновременно быстрых и эффективных методов оптимизации. В других случаях необходимо оценивать ряд параметров передачи, опираясь лишь на косвенные признаки, для чего отлично подходят методы машинного обучения. В рамках данного семинара мы рассмотрим различные математические задачи, встречающиеся при построении базовой станции пятого поколения: многоантенную передачу (massive multiple input multiple output), оптимальный выбор пользователей (user-pairing) и адаптивную подстройку параметров передачи (link-adaptation).	Björnson, Emil, Jakob Hoydis, and Luca Sanguinetti. "Massive MIMO networks: Spectral, energy, and hardware efficiency." Foundations and Trends in Signal Processing 11.3-4 (2017): 154-655. Tse, David, and Pramod Viswanath. Fundamentals of wireless communication. Cambridge university press, 2005. Zaidi, Ali, et al. 5G Physical Layer: principles, models and technology components. Academic Press, 2018. Видео
01.04.2022	Нет спецсеминара	
08.04.2022	Никита Гущин, ШАД Autoformer and Autoregressive Denoising Diffusion Models for Time Series Forecasting В докладе представлен разбор двух статей посвящённых предсказанию временных рядов. В "Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting" (NIPS, 2021) упор делается на использование зарекомендовавших себя трансформеров для предсказания временных рядов на как можно большее число шагов вперёд, для чего авторы предложили новый эффективный (не квадратичный по асимптотике) вариант на замену self-attention блоку и использование разложения временного ряда на тренд и сезонность внутри декодера. В "Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting" (ICML 2021), основной мотиваций является получение модели для вероятностного предсказания многомерных временных рядов, для того чтобы иметь возможность оценить неопределённость получаемых предсказаний. Авторы предлагают это сделать путём обуславливания диффузионной модели на hidden state рекуррентной нейронной сети.	Презентация Статья 1, Статья 2 Видео
15.04.2022	Иван Щекотов, Samsung AI Center Bilateral Denoising Diffusion Models for Fast and High-Quality Speech Synthesis Основная проблема, с которой сталкиваются диффузионные модели – это медленная скорость сэмплирования. В основном, существующие подходы, ускоряющие процессы сэмплирования, зависят от детерминированного расписания шумов. При этом его оптимальность варьируется на различных датасетах. В этом докладе основной упор будет сделан на разборе модели с ICLR 2022 “BDDM: Bilateral Denoising Diffusion Models for Fast and High-Quality Speech Synthesis”, в которой авторы значительно ускоряют процесс сэмплирования благодаря репараметризации шумов для прямого процесса с помощью schedule network. Благодаря связи с DDPM процедура обучения score network остается неизменной, в то время как предсказание шума позволяет строить более короткие диффузионные цепочки для инференса. В отличие от параллельной работы Variational Diffusion Models, где SNR используется для noise scheduling, предсказание шума зависит не только от шага времени, но и от зашумленного сэмпла в данный момент времени. Выводятся теоретические предпосылки, которые доказывают корректность процедур обучения и	Статья

	сэмплирования BDDM, такие как: - верхние границы на предсказываемый шум - эквивалентность между обучением нижней вариационной оценки для DDPM и BDDM для score network - функция потерь для обучения schedule network при данной оптимальной score network.	
22.04.2022	Дмитрий Курцев, МГУ им. М.В. Ломоносова MLP архитектура для решения задач CV и NLP Трансформеры стали одним из самых важных архитектурных открытий в области deep learning и позволили добиться многих прорывов за последние несколько лет в задачах NLP и CV. В данном докладе представляются простые архитектуры, основанные на многослойном перцептроне, MLP-Mixer и gMLP. Они ставят под сомнение необходимость слоя self-attention для достижения хорошей точности. MLP сети получают конкурентоспособные результаты в задачах классификации текстов и изображений, при этом затраты на pre-train и fine-tuning сопоставимы с sota.	Презентация Видео
29.04.2022	Олег Дешеулин, НИУ ВШЭ, Сколтех Дифференцирумая реконструкция трехмерных сцен по снимкам В последние годы набирает всё больше популярности тема синтеза фото сцен с новых ракурсов на основе уже известных (novel view synthesis). Революцию в этой области произвело решение Neural Radience Fields (NeRF). Поговорим про то как оно устроено, как его улучшают с помощью эффективных и кэшируемых репрезентаций и как можно улучшить сэмплирование в подсчете интеграла для определения цвета пикселя изображения.	Презентация Видео
06.05.2022	Андрей Охотин, МГУ им. М.В. Ломоносова Минимизация энергии в моделях Изинга с помощью нейронных сетей В противоположность задачам оптимизации с непрерывными переменными, переменные в задачах дискретной оптимизации принимают только дискретные значения. Одной из таких задач является поиск минимума субмодулярного функционала энергии в модели Изинга. Мы рассмотрим подходы построения универсальных солверов этой задачи с применением нейронных сетей. Поговорим о проблемах, возникающих в постановках, допускающих генерацию обучающей выборки, а также о способе обучения нейронных сетей на сложный для оптимизации функционал качества.	Презентация Видео
13.05.2022	Данила Дорошин, Huawei Предыскажение сигналов усилителей Усилитель мощности является важной частью базовой станции мобильной связи 5G. Кроме непосредственно усиления сигнала, усилитель генерирует нежелательный шум, вызванный нелинейностью в работе транзисторов. На выходе усилителя получается электромагнитная волна большой мощности, поэтому очистка выходного сигнала цифровыми алгоритмами не представляется возможной. Вместо очистки выходного сигнала используется подход цифрового предыскажения, заключающийся в предварительной обработке сигнала специальной функцией. В случае достаточно точного приближения данной функции к обратной функции усилителя такая технология позволяет понизить уровень шума до приемлемых показателей. Самыми простыми в реализации и самыми распространенными в литературе являются полиномиальные модели. Они линейны по коэффициентам, поэтому их оптимизация, как правило, является выпуклой задачей. Однако, в случае сигналов 5G полиномиальные модели могут не дотягивать до требуемой	Презентация Видео

	точности, либо могут потребоваться полиномы высокого порядка (больше 10) и от большого количества входных отсчётов (больше 30). В качестве более выразительных моделей обычно рассматриваются модели типа Wiener-Hammerstein (стр. 11). Фактически модель Wiener-Hammerstein является нейронной сетью из двух слоёв: первый слой свёрточный, второй – полиномиальный. Такая модель уже перестаёт быть линейной по коэффициентам и требует привлечения более сложных методов оптимизации. Вычислительная сложность нейронных сетей, как правило, исчисляется в более чем сотнями тысяч умножений и таком же количестве сложений. Однако, индустрия беспроводной связи выдвигает довольно жесткие требования к применяющимся моделям, ограничивая вычисления порядком 1000 умножений на прямом проходе модели. Данная особенность заставляет искать компактные архитектуры моделей. Вторая особенность состоит в необходимости реализации модели в фиксированной арифметике. Стоит отметить необходимость постоянно адаптировать модель к меняющейся окружающей среде, типу входного сигнала, постепенному износу усилителя. Адаптация происходит за счет постоянного захвата выходного сигнала с усилителя. При этом метод адаптации такой как обратное распространение ошибки также должен работать в фиксированной арифметике.	
20.05.2022	Александр Коротин, Сколтех, AIRI Neural Optimal Transport Solving optimal transport (OT) problems with neural networks has become widespread in machine learning. The majority of existing methods compute the OT cost and use it as the loss function to update the generator in generative models (Wasserstein GANs). In this presentation, I will discuss the absolutely different and recently appeared direction - methods to compute the OT plan (map) and use it as the generative model itself. Recent advances in this field demonstrate that they provide comparable performance to WGANs. At the same time, these methods have a wide range of superior theoretical and practical properties. The presentation will be mainly based on our recent pre-print "Neural Optimal Transport" https://arxiv.org/abs/2201.12220. I am going to present a neural algorithm to compute OT plans (maps) for weak & strong transport costs. For this, I will discuss important theoretical properties of the duality of OT problems that make it possible to develop efficient practical learning algorithms. Besides, I will prove that neural networks actually can approximate transport maps between probability distributions arbitrarily well. Practically, I will demonstrate the performance of the algorithm on the problems of unpaired image-to-image style transfer and image super-resolution.	Презентация Related Work Neural Optimal Transport https://arxiv.org/pdf/2201.12220.pdf Unpaired Image Super-Resolution with Optimal Transport Maps https://arxiv.org/pdf/2202.01116.pdf Generative Modeling with Optimal Transport Maps https://openreview.net/pdf?id=5JdLZg346Lw Do Neural Optimal Transport Solvers Work? A Continuous Wasserstein-2 Benchmark https://proceedings.neurips.cc/paper/2021/file/7a6a6127ff85640ec69691fb0f7cb1a2-Paper.pdf Видео

Осенний семестр 2021 г.

Дата	Докладчик и тема	Материалы
17.09.2021	Надежда Чиркова, НИУ ВШЭ Neural Program Synthesis (Часть 1)	Видео Презентация

	Аннотация: Языковые модели наподобие BERT и GPT-3 достигли высоких результатов во многих прикладных задачах, включая машинный перевод, генерацию текста и информационный поиск, и уже вышли за пределы обработки текста. Летом 2021 года компания OpenAI представила модель Codex, способную генерировать качественный программный код по описанию на естественном языке и контексту и встроенную в качестве ассистента программиста GitHub Copilot в редактор Visual Studio. Как и GPT-3, модель Codex была предобучена на огромном наборе данных, в данном случае содержащем исходный код, и далее дообучена на задачу генерации кода по текстовому описанию.	Статья 1, 2, 3
24.09.2021	Сергей Трошин, НИУ ВШЭ Neural Program Synthesis (Часть 2) Аннотация: На семинаре мы рассмотрим особенности языковых моделей для задачи генерации кода, в частности поговорим про Codex, но затроним и смежные исследования. В докладе будет уделено внимание анализу ошибок языковых моделей, способам семплирования из них, рассмотрены варианты улучшения качества языковых моделей: с помощью дообучения, взаимодействия с пользователем, и др.	Видео Презентация Статья 1, 2, 3
01.10.2021	Пётр Мокров, Сколтех, МФТИ Large-Scale Wasserstein Gradient Flows Абстракт: Wasserstein gradient flows provide a powerful means of understanding and solving many diffusion equations. Specifically, Fokker-Planck equations, which model the diffusion of probability measures, can be understood as gradient descent over entropy functionals in Wasserstein space. This equivalence, introduced by Jordan, Kinderlehrer and Otto, inspired the so-called JKO scheme to approximate these diffusion processes via an implicit discretization of the gradient flow in Wasserstein space. Solving the optimization problem associated to each JKO step, however, presents serious computational challenges. We introduce a scalable method to approximate Wasserstein gradient flows, targeted to machine learning applications. Our approach relies on input-convex neural networks (ICNNs) to discretize the JKO steps, which can be optimized by stochastic gradient descent. Unlike previous work, our method does not require domain discretization or particle simulation. As a result, we can sample from the measure at each time step of the diffusion and compute its probability density. We demonstrate our algorithm's performance by computing diffusions following the Fokker-Planck equation and apply it to unnormalized density sampling as well as nonlinear filtering.	Видео Презентация Препринт Статья 1, 2, 3, 4
08.10.2021	Григорий Бартош, JetBrains Research Diffusion models (часть1) Абстракт: Вероятно, вы уже что-то слышали про диффузные модели и видели провокационные названия статей типа "Diffusion Models Beat GANs on Image Synthesis". Диффузные модели это относительно новый класс моделей, которые показывают sota результаты в задачах оценки плотности распределения и генерации данных. На некоторых задачах генерации изображений они показывают результаты лучше, чем GAN'ы. Мы разберем стандартные диффузные модели и посмотрим на некоторые более поздние обобщения. Поймем, как обучать эти модели, оценивать с их помощью плотность распределения, генерировать данные (априорно и условно) и варьировать объем используемых вычислительных ресурсов во время инференса.	Видео Презентация Статья 1, 2, 3, 4
15.10.2021	Виктор Оганесян, НИУ ВШЭ	Видео

	Diffusion models (часть 2) Абстракт: на данном семинаре мы рассмотрим непрерывное обобщение моделей, рассмотренных на предыдущем семинаре с помощью стохастических дифференциальных уравнений. Такое обобщение оказалось очень эффективным и побило предыдущие модели на нескольких датасетах. Мы рассмотрим как оно включает в себя старые модели, позволяет создавать новые и какие расширения позволяет делать благодаря непрерывности.	Статья
22.10.2021	Виктор Оганесян, НИУ ВШЭ Diffusion models (часть 3)	Видео
29.10.2021	Efstratios Gavves, University of Amsterdam The Machine Learning of Time and Dynamics … with an Outlook towards the Sciences Абстракт: In the past decades, the impressive progress in machine learning and applications -like computer vision- was mainly by assuming (or enforcing) that data is static and usually of spatial-only nature, that data is i.i.d, that learning correlations suffices for high predictive accuracies. In the real world, however, data and processes are typically (spatio-) temporal, dynamic, non-stationary, non-iid, causal. This leads to paradoxical situations for learning algorithms. In this talk, I will first present my vision for a new type of learning that embraces temporality and dynamics. I will then discuss recent work that connects complexity in deep stochastic models, like hierarchical VAEs, with phase transitions, pointing perhaps to a link to statistical physics. I will continue with discussing how simple ways of introducing roto-translation equivariance can greatly improve standard neural relational inference in modelling dynamics of complex interacting dynamical systems. Last, I will present our latest attempts in scaling up causal discovery by at least two orders of magnitude compared to the recent literature. I will close with drawing a connection between machine learning and the sciences, whose interface -I believe- is deeply temporal and dynamical, and will inspire the great next breakthroughs.	Презентация Видео
05.11.2021	Нет семинара (праздничный день)	
12.11.2021	Михаил Фигурнов, DeepMind Highly accurate protein structure prediction with AlphaFold Predicting a protein's structure from its primary sequence has been a grand challenge in biology for the past 50 years, holding the promise to bridge the gap between the pace of genomics discovery and resulting structural characterization. In this talk, we will describe work at DeepMind to develop AlphaFold, a new deep learning-based system for structure prediction that achieves high accuracy across a wide range of targets. We demonstrated our system in the 14th biennial Critical Assessment of Protein Structure Prediction (CASP14) across a wide range of difficult targets, where the assessors judged our predictions to be at an accuracy "competitive with experiment" for approximately 2/3rds of proteins. The talk will focus on the underlying machine learning ideas, while also touching on the implications for biological research.	Презентация Статья Видео
19.11.2021	Нет семинара	

26.11.2021	Вадим Титов, МФТИ Controlling GANs Latent Space Modern GAN architectures generate highly realistic images in a variety of domains. Much recent works have focused on understanding how its latent space is connected with generated image semantic. It is discovered that there exist meaningful latent manipulations that allow to semantically edit image. Many proposed methods include supervision from pretrained models which is a strong limitation. This weakness could be eliminated by unsupervised methods that have their own disadvantages. In this talk we will describe the main directions (supervised, unsupervised, text-guided) and current state-of-the-art methods of semantic image manipulation through GAN latent space.	Презентация Полезные ссылки: Common ideas on latent space manipulations: https://arxiv.org/abs/1907.10786 Key paper on unsupervised approach: https://arxiv.org/abs/2002.03754 StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery: https://arxiv.org/pdf/2103.17249 Видео
03.12.2021	Prof. Mikhail (Misha) Belkin, University of California, San Diego From classical statistics to modern deep learning Recent empirical successes of deep learning have exposed significant gaps in our fundamental understanding of learning and optimization mechanisms. Modern best practices for model selection are in direct contradiction to the methodologies suggested by classical analyses. Similarly, the efficiency of SGD-based local methods used in training modern models, appeared at odds with the standard intuitions on optimization. I will present evidence, empirical and mathematical, that necessitates revisiting classical statistical notions, such as over-fitting. I will continue to discuss the emerging understanding of generalization, and, in particular, the "double descent" risk curve, which extends the classical U-shaped generalization curve beyond the point of interpolation. While our understanding has significantly grown in the last few years, a key piece of the puzzle remains -- how does optimization align with statistics to form the complete mathematical picture of modern ML?	Видео
10.12.2021	Нет семинара	
17.12.2021	Роман Суворов, Advanced Image Manipulation Lab @ Samsung AI Center Moscow Resolution-robust Large Mask Inpainting with Fourier Convolutions (and a brief survey of image inpainting) I will present our recent paper - Resolution-robust Large Mask Inpainting with Fourier Convolutions - which was accepted to WACV'22. Image inpainting is the problem of re-generating the missing areas in images. While significant progress has been achieved in the field since adoption of deep generative networks in approximately 2016, there is still a large room for improvement. In the paper we present a new system - LaMa - which addresses some of the most noticeable limitations of the existing approaches. LaMa is able to re-generate repetitive and structured backgrounds (but not limited to) and can do it in high resolution, while being trained using only 256x256	Статья 1, 2 Видео

	images. This is possible thanks to the fast Fourier convolutions in the generator, a high receptive field perceptual loss and large training masks. I will also briefly review related work and discuss how LaMa compares to state of the art approaches.	

Весенний семестр 2021 г.

Дата	Докладчик и тема	Материалы
26.02.2021	Алексей Наумов, НИУ ВШЭ Случайные матрицы: теория и приложения Теория случайных матриц и методы, используемые при исследовании случайных матриц, играют важную роль в различных разделах теоретической и прикладной математики. Случайные матрицы возникли из приложений, сначала в анализе данных, а позже в качестве статистических моделей в квантовой механике, вычислительной математике, финансовой инженерии, теории информации, машинном обучении и других областях. В последние двадцать лет произошел настоящий бум в развитии теории случайных матриц. Были получены прорывные результаты. В своем докладе я расскажу об основных законах, возникающих в поведении спектра случайных матриц, а также о некоторых приложениях. Доклад частично основан на моих совместных работах с Фридрихом Гётце и Александром Тихомировым.	Презентация Видео
05.03.2021	Денис Ракитин, НИУ ВШЭ Stochastic gradient estimation for discrete variables Training models with discrete latent variables remains a challenging task because of difficulties in accurately estimating the gradients. To address this problem, one can use reparameterizable relaxations (e.g. Gumbel-Softmax), which usually give a low-variance estimate. However, this leads to biased gradients and puts additional constraints on the model. Alternatively, estimators built on score-function methods are unbiased, more general, but require a careful design of variance reduction techniques. In this talk we will discuss a family of recently proposed methods for categorical variables that make use of variable augmentation, REINFORCE and Rao-Blackwellization. We will also cover the analysis of the	Презентация Статья 1, 2, 3, 4

	REINFORCE estimator with a leave-one-out baseline, which was shown to be effective despite its simplicity.	
12.03.2021	Александр Гришин, SAIC Moscow Meta-learning in neural networks В докладе будет рассказано о мета-обучении в глубоком обучении [1]. В начале мы рассмотрим высокоуровневую классификацию современных алгоритмов мета-обучения. В частности, про то, что обучается, как обучается и с какой целью. Далее мы более подробно остановимся на мета-обучении в обучении с подкреплением (RL) и рассмотрим ключевые различия подходов в этой области. В конце мы рассмотрим несколько конкретных примеров применения [2-4], начиная от простейшего обучения награды и заканчивая мета-обучением самих RL алгоритмов.	Презентация Статьи 1, 2, 3, 4
19.03.2021	Михаил Архипов Complete Likelihood Objective for Latent Variable Models Existing probabilistic approaches to learning deep latent variable models rely on marginal likelihood objective, while optimization of complete likelihood (CL) is left for fully observed cases. In this work, we show how to optimize CL in the latent variable setting. We treat a sample from prior distribution as a set of target latent variables with unknown permutation. This step allows to replece approximate inference with combinatorial optimization. In contrast to variational approaches, CoLike does not experience posterior collapse and learns more informative latents. Furthermore, due to absence of an encoder CoLike does not need special techniques for learning discrete latent variable models. Finally, we show that CoLike bridges optimal transport and probabilistic frameworks.	Презентация Статьи: Bojanowski P., Joulin A. Unsupervised learning by predicting noise //International Conference on Machine Learning. – PMLR, 2017. – С. 517-526. Patrini G. et al. Sinkhorn autoencoders //Uncertainty in Artificial Intelligence. – PMLR, 2020. – С. 733-743. Hsu D., Shi K., Sun X. Linear regression without correspondence //arXiv preprint arXiv:1705.07048. – 2017.
26.03.2021	Евгений Голиков, École polytechnique fédérale de Lausanne Former affiliation: DeepPavlov.ai, Moscow Institute of Physics and Technology Tensor Programs-1 We shall discuss a formalism of Tensor programs that allows one to express neural network computation (e.g. forward and backward passes) for a wide class of neural nets. The formalism is equipped with a theorem (the Master theorem) that reasons about the distributions of random variables of the program in the limit of infinite width. Several previous results about infinite width nets: i.e. convergence to a Gaussian process at initialization and convergence of a neural tangent kernel to a deterministic variable, can be deduced as simple corollaries of the Master theorem.	Презентация Статьи 1, 2
02.04.2021	Полина Кириченко Continual learning in neural networks: on catastrophic forgetting and beyond Learning new tasks continually without forgetting on a constantly changing data distribution is essential for real-world problems but is challenging for modern deep learning. Deep learning models suffer from catastrophic forgetting: when presented with a sequence of tasks, deep neural networks can successfully learn the new tasks, but the performance on the old tasks degrades. In this talk, I will present an overview of the continual learning algorithms including well-established methods as well as recent state-of-the-art approaches. We will talk about several continual learning scenarios (task-, class-, and domain-incremental learning), review the most common approaches in alleviating forgetting and discuss other	Презентация Статьи: Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks." Proceedings of the national academy of sciences 114.13 (2017): 3521-3526. Parisi, German I., et al. "Continual lifelong learning with neural networks: A review." Neural Networks 113 (2019):

	challenges in the field beyond catastrophic forgetting (including forward & backward transfer, learning on continuously drifting data and continual learning of unsupervised tasks).	54-71. Hadsell, Raia, et al. "Embracing Change: Continual Learning in Deep Neural Networks." Trends in Cognitive Sciences (2020). Hsu, Yen-Chang, et al. "Re-evaluating continual learning scenarios: A categorization and case for strong baselines." arXiv preprint arXiv:1810.12488 (2018). Van de Ven, Gido M., and Andreas S. Tolias. "Three scenarios for continual learning." arXiv preprint arXiv:1904.07734 (2019). Видео
09.04.2021	Евгений Голиков,École polytechnique fédérale de Lausanne Former affiliation: DeepPavlov.ai, Moscow Institute of Physics and Technology Tensor Programs-2	Презентация Видео
16.04.2021	Евгений Егоров Learning differential equations that are easy to solve Модель Neural ODE сопоставляет наблюдаемыми данными векторное поле. Вне окрестностей обучающей выборки модель имеет произвол. Естественный способ борьбы с этим недостатком применение какой-либо регуляризации для выбора решения. В докладе рассматривается предложенная (Jacob Kelly et al, 2020) идея измерения “простоты” решения с помощью нормы K-1-й производной правой части и связанной с этой идеей технический инструментарий.	Презентация Статья Видео
23.04.2021	Айбек Аланов, Олег Иванов (SAIC Moscow) Audio Synthesis and Bandwidth Extension Абстракт: В докладе будет рассказано о задаче генерации звука с помощью нейросетей, и в частности мы рассмотрим задачу увеличения разрешения звука с помощью условных генеративных моделей. В начале мы сделаем краткое введение в область обработки сигналов. Дальше рассмотрим стандартные генеративные модели для звука, основанные на нейросетях, и последние достижения в этой области. В конце посмотрим, как работают эти модели на конкретных датасетах.	Презентация Статья 1, 2, 3, 4 Видео
30.04.2021	Андрей Чернов, НИУ ВШЭ Minibatch acceptance test for Metropolis-Hastings Абстракт: Для выполнения Metropolis-Hastings теста необходимо использовать всю выборку, что сдерживает масштабирование методов MCMC на задачи с большими данными. В докладе рассматривается идея (Daniel Selta et al, 2017), позволяющая провести MH тест по батчу, не используя всю выборку. Основное внимание будет сконцентрировано на недостатках предложенного метода и будут предложены способы улучшения данного подхода.	Презентация Статья

07.05.2021	Павел Измаилов, New York University What Are Bayesian Neural Network Posteriors Really Like?	Презентация Видео
14.05.2021	Артём Цыпин, МГУ, Samsung AI Center Imitation Learning from Observations Во многих задачах в обучении с подкреплением ключевую роль играет проектирование функции награды. Однако, для многих задач гораздо легче предоставить демонстрации требуемого поведения. Задачу обучения агента с экспертных демонстраций, в которых не содержатся действия, решают с помощью имитационного обучения с наблюдений. В докладе будут рассмотрены методы для имитационного обучения с наблюдений, а также предложен метод, основанный на оптимизации точечной взаимной информации.	Презентация Статьи 1, 2, 3, 4, 5 Видео
21.05.2021	Фёдор Лебедь, МГУ Дифференцирование через решение оптимизиационных задач для настройки гиперпараметров В большинстве моделей машинного обучения присутствуют так называемые гипер-параметры. В отличие от параметров, которые настраиваются на обучающую выборку, эти гипер-параметры подбираются таким образом, чтобы модель, настроенная на обучающую выборку, вела себя наилучшим образом на тестовой или валидационной выборке. В таком случае настройка гипер-параметров обычно выполняется перебором по сетке. Такой подход, однако, неприменим при большом количестве гипер-параметров в силу проклятия размерности. В рамках данного доклада будут детально рассмотрены два метода вычисления градиентов по гипер-параметрам для градиентной оптимизации последних.	Презентация Статьи 1, 2, 3

Осенний семестр 2020 г.

Дата	Докладчик и тема	Материалы
18.09.2020	Иван Назаров, ADASE Skoltech Bayesian Sparsification Methods for Deep Complex-valued Networks Доклад посвящен обобщению Вар-Дропаута для комплексных нейросетей. Будет кратко рассказано про отличия комплексных от обычных нейросетей, представлены сравнения с вещественным Вар-Дропаутом и приведены результаты сжатия глубокой сетки в задаче аннотации музыки	Презентация Статья 1 Видео

25.09.2020	Нет спецсеминара	
02.10.2020	Ермек Капушев, Skoltech, ADASE group Random Fourier Features based on Quadratures and their use for a) density estimation b) SLAM (Simultaneous Localization and Mapping for Robotics) based on GP model Доклад будет посвящен аппроксимации ядровых функция с помощью случайных признаков. Будет рассказано про применение квадратурных правил для генерации случайных признаков, а также будет рассмотрены примеры применения случайных признаков в задаче оценки плотности и в задаче одновременной локализации и построения карты	Презентация Статья Видео
09.10.2020	Кирилл Струминский, НИУ ВШЭ Обобщение Гумбель-софтмакс трика Gumbel Softmax Trick предложил эффективный и простой в реализации метод обучения глубоких моделей с категориальными скрытыми переменными. Однако его время работы линейно по размеру носителя скрытой переменной. Поэтому, если скрытая переменная соответствует дереву разбора входного предложения или перестановке элементов входной последовательности, Gumbel Softmax Trick оказывается неприменим на практике. В докладе речь пойдет о ряде недавно предложенных подходов для обобщения метода на случай структурных скрытых переменных	Презентация Статья 1 Статья 2 Статья 3 Видео
16.10.2020	Екатерина Лобачева, НИУ ВШЭ On Power Laws in Deep Ensembles Ансамбли нейронных сетей широко применяются на практике, особенно для задач, в которых важна устойчивая оценка неопределенности модели. В докладе мы посмотрим как ведет себя качество ансамбля в терминах NLL/CNLL как функция от количества сетей в ансамбле, размера этих сетей и общего числа параметров в модели. Мы увидим, что во многих случаях качество ведет себя как степенной закон, что само по себе любопытно, плюс позволяет предсказывать возможную прибавку в качестве при увеличении моделей. Также мы рассмотрим случай фиксированного бюджета по памяти и поймем как лучше его распределять - брать мало больших сетей или много маленьких - и как это распределение предсказывать для конкретных задач с помощью обнаруженных степенных законов	Презентация Статья Видео
23.10.2020	Евгений Голиков, DeepPavlov.ai, Neural Networks and Deep Learning lab., MIPT Infinitely wide nets Many problems in theoretical understanding of neural nets come from the fact that it is hard to reason about their training dynamics. In particular, one cannot generally guarantee global convergence of gradient descent --- a fact typically observed for realistic networks and data. Moreover, all generalization bounds that do not take the training dynamics into account turn out to be vacuous.	Презентация Видео

	Fortunately, the training dynamics of neural nets substantially simplifies in the limit of infinite width. One of the limit, the NTK limit, is driven by a constant kernel which can be estimated via Monte-Carlo. We shall discuss how this limit can be used to obtain optimization and generalization guarantees for sufficiently wide networks. Another limit, the mean-field limit, leads to a quantitatively different limit model. The reason why we have two different limits is the difference in hyperparameter scaling with width. We shall show how different hyperparameter scalings result in different limit models, and discuss which limit model should be a better proxy for realistic finite-width nets.	
30.10.2020	Владимир Сурдин Астрономия	Видео
06.11.2020	Дмитрий Копитков General Probabilistic Surface Optimization Probabilistic inference, such as density (ratio) estimation, is a fundamental and highly important problem that needs to be solved in many different domains including robotics and computer science. Recently, a lot of research was done to solve it by producing various objective functions optimized over neural network (NN) models. Such Deep Learning (DL) based approaches include unnormalized and energy models, as well as critics of Generative Adversarial Networks, where DL has shown top approximation performance. In this research we contribute a novel algorithm family, which generalizes all above, and allows us to infer different statistical modalities (e.g. data likelihood and ratio between densities) from data samples. The proposed unsupervised technique, named Probabilistic Surface Optimization (PSO), views a model as a flexible surface which can be pushed according to loss-specific virtual stochastic forces, where a dynamical equilibrium is achieved when the pointwise forces on the surface become equal. Concretely, the surface is pushed up and down at points sampled from two different distributions, with overall up and down forces becoming functions of these two distribution densities and of force intensity magnitudes defined by the loss of a particular PSO instance. Upon convergence, the force equilibrium associated with the Euler-Lagrange equation of the loss enforces an optimized model to be equal to various statistical functions, such as data density, depending on the used magnitude functions. Furthermore, this dynamical-statistical equilibrium is extremely intuitive and useful, providing many implications and possible usages in probabilistic inference. We connect PSO to numerous existing statistical works which are also PSO instances, and derive new PSO-based inference methods as demonstration of PSO exceptional usability. Additionally, we investigate the impact of Neural Tangent Kernel (NTK) on PSO equilibrium. Our study of NTK dynamics during the learning process emphasizes the importance of the model kernel adaptation to the specific target function for a good learning approximation.	Презентация Статья 1; Статья 2; Статья 3; Видео
13.11.2020	Надежда Чиркова, НИУ ВШЭ Adapting Natural Language Processing to Source Code Processing: Handling Syntactic Structure and Identifiers Initially developed for natural language processing, Transformers and RNNs are now widely used for source code processing, due to the format similarity between source code and text. In contrast to natural language, source code is strictly structured, i. e. follows the syntax of the programming language.	Презентация Статья 1; Статья 2; Статья 3;

	Another important property of source code is invariance to renaming user-defined identifiers. I will tell you about our research on utilizing both mentioned properties in Transformer and recurrent architectures. I will first describe our empirical study on the capabilities of Transformers to utilize syntactic information, including the comparison of several recently proposed tree-processing Transformer mechanisms on three code processing tasks (code completion, function naming, and bug fixing), and testing Transformers in a so-called anonymized setting, in which all variables are replaced with unique placeholders. Secondly, we will discuss the practical applicability of the mentioned anonymized setting. Thirdly, I will present our dynamic embedding architecture for processing anonymized variables in the RNNs.	Видео
20.11.2020	Защита Кирилла Неклюдова Тема: Байесовский подход в глубинном обучении: улучшение дискриминативных и генеративных моделей Ссылка приглашения https://zoom.us/j/97098090066 Идентификатор конференции 970 9809 0066	
27.11.2020	Максим Кодрян, НИУ ВШЭ Double Descent, flat minima, and SGD The Double Descent (DD) phenomenon has recently appeared as a particularly intriguing finding in the Deep Learning community. While most works tackle the famous model-wise DD (test risk vs. model size) from both empirical and theoretical points of view, much less attention is paid to the no less mystifying epoch-wise DD effect (test risk vs. number of training epochs). Another interesting observation, gaining momentum in the most recent studies, is the conventional "flat minima" argument: the wider the minimum the better it generalizes. In this talk, we will try to link the epoch-wise Double Descent with model dynamics on the loss surface: the model enjoys the second test risk descent exactly when it traverses from the firstly found sharp unstable regions to flat well-generalizing minima. We will also regard the implicit regularization of Stochastic Gradient Descent (SGD), aiding neural networks to converge into such wide "uniform" optima.	Презентация Статья 1; Статья 2; Статья 3; Видео
04.12.2020	Артем Гадецкий, НИУ ВШЭ Differentiation through solutions to optimization problems In this talk we will discuss general methodology for embedding solutions to parametrized constrained convex optimization as layers for deep neural networks (DNNs). Particular examples include but not limited to parametrized Quadratic Programs (QP) as well Disciplined Parametrized Programming (DPP), framework which allows bypassing error-prone process of manual convertation of optimization problems to canonical forms that greatly accelerates prototyping and application to DNNs.	Презентация Статья 1; Статья 2; Статья 3; Видео
18.12.2020	Александр Коротин, Skolkovo Institute of Science and Technology Wasserstein-2 Generative Networks	Презентация Статья 1;

	We propose a novel end-to-end non-minimax algorithm for training optimal transport mappings for the quadratic cost (Wasserstein-2 distance). The algorithm uses input convex neural networks and a cycle-consistency regularization to approximate Wasserstein-2 distance. In contrast to popular entropic and quadratic regularizers, cycle-consistency does not introduce bias and scales well to high dimensions. From the theoretical side, we estimate the properties of the generative mapping fitted by our algorithm. From the practical side, we evaluate our algorithm on a wide range of tasks: image-to-image color transfer, latent space optimal transport, image-to-image style transfer, and domain adaptation.	Статья 2; Статья 3; Статья 4; Видео

Весенний семестр 2020 г.

Дата	**Докладчик и тема**	Материалы
28.02.2020	Александр Лыжов, Samsung AI Center Moscow, Research Scientist Model calibration In many real-world applications we would like the probabilities that the model outputs (e.g. class probabilities in classification) to be correct in some sense (e.g. to match the actual probabilities of class occurrence). This property of models is called calibration. In this talk I will first do a introduction to various aspects of calibration: definitions of calibration errors, estimators of these errors, calibration of neural networks. Then I will talk about developments that occured in understanding of calibration in 2019 in depth. I want to focus on unbiased calibration estimators and hypothesis testing for calibration in particular. If we have time after that, we may talk about calibration of regression and differentiable calibration losses in neural network training.	Calibration tests in multi-class classification: A unifying framework On Calibration of Modern Neural Networks Trainable Calibration Measures for Neural Networks from Kernel Mean Embeddings Упомянутые в докладе статьи: Beyond temperature scaling: Obtaining well-calibrated multiclass probabilities with Dirichlet calibration Learn-By-Calibrating: Using Calibration as a Training Objective Greedy Policy Search: A Simple Baseline for Learnable Test-Time Augmentation Pitfalls of In-Domain Uncertainty Estimation and Ensembling in Deep Learning Презентация Видео
06.03.2020	Денис Ракитин, НИУ ВШЭ Neural Program Synthesis	Видео Презентация

	The problem of program learning consists of generating a computer program consistent with some specification. Contemporary studies showed that using neural program synthesis approach for solving this type of task can result in a more robust and less data dependent algorithm comparing to the classical methods. Moreover, adding a program synthesis module as a building block of another neural-based algorithm can provide it with prior structural knowledge and make its performance more interpretable. In this talk I will make an introduction to this approach provided with some examples of usage. Main part of the speech will be in covering 2 recent papers that successfully combine neural program synthesis with deep representation learning in application to visual question answering problem.	Статья 1 ; Статья 2
13.03.2020	Андрей Малинин, Яндекс Uncertainty in Structured Prediction Uncertainty estimation is important for ensuring safety and robustness of AI systems, especially for high-risk applications. While much progress has recently been made in this area, most research has focused on un-structured prediction, such as image classification and regression tasks. However, while task-specific forms of confidence score estimation have been investigated by the speech and machine translation communities, limited work has investigated general uncertainty estimation approaches for structured prediction. Thus, this work aims to investigate uncertainty estimation for structured prediction tasks within a single unified and interpretable probabilistic ensemble-based framework. We consider uncertainty estimation for sequence data at the token-level and complete sequence-level, provide interpretations for, and applications of, various measures of uncertainty and discuss the challenges associated with obtaining them. This work also explores the practical challenges associated with obtaining uncertainty estimates for structured predictions tasks and provides baselines for token-level error detection, sequence-level prediction rejection, and sequence-level out-of-domain input detection using ensembles of auto-regressive transformer models trained on the WMT'14 English-French and WMT'17 English-German translation and LibriSpeech speech recognition datasets.	Видео Презентация Статья
20.03.2020	Екатерина Лобачева, научный сотрудник лаборатории компании Самсунг-НИУ ВШЭ BERT and his friends (отменен)	
27.03.2020	Александра Волохова, стажёр-исследователь лаборатории компании Самсунг-НИУ ВШЭ Neural Stochastic Differential Equations Neural SDE расширяет модель neural ODE с помощью введения стохастичности в систему дифференциальных уравнений. На семинаре мы обсудим, как и зачем добавлять стохастичность в neural ODE и для решения каких прикладных задач это может быть полезно. В процессе обсуждения рассмотрим подходы, предложенные в следующих статьях: 1, 2, 3, 4 Ссылка для участия в видеоконференции Zoom: https://zoom.us/j/759154498	Презентация Статьи: 1, 2, 3, 4
03.04.2020	Александр Фрицлер, Samsung AI Center Moscow, Research Scientist Quantization of neural networks	Видео Презентация

	Аннотация: Квантизация сетей - это процесс их представления с помощью чисел с низкой точностью. Такое представление необходимо для уменьшения объёма памяти, необходимого для хранения весов, а также для применения сетей на специальных процессорах под названием Neural Processor Unit. Мы поговорим о том, как можно уже обученную сеть представить в таком виде, а также, как обучить такую сеть с нуля с использованием Pytorch	Статьи 1, 2, 3 Конференция
10.04.2020	Максим Кочуров, Samsung AI Center Moscow, Research Scientist Hyperbolic Deep Learning Hyperbolic Deep Learning gained attention due to its ability to work with and represent hierarchical relations. However, we do not yet have enough tools to work in non-Euclidean space. Several works present proof of concept results on various tasks: word embeddings, text classification, node classification, link prediction, and others. Methods discussed include Hyperbolic GloVe, GRU, VAE, graph embeddings, and graph neural networks. These works introduce new concepts and link Euclidean models to their Hyperbolic extensions. While having fairly simple baselines, they provide some evidence where Hyperbolic geometry might be more suitable. During the talk, we'll try to answer the following essential question: when do we need Hyperbolic geometry in deep learning?	Видео Презентация Статьи 1, 2, 3, 4, 5, 6
17.04.2020	Екатерина Лобачева, научный сотрудник лаборатории компании Самсунг-НИУ ВШЭ BERT: model, analysis and modifications В современном NLP при решении многих задач используются контекстуальные эмбеддинги, предобученные на большом объеме неразмеченных данных. В данном докладе мы поговорим о том, что такое контекстуальные эмбеддинги, и обсудим подробно наиболее базовую и часто используемую модель - BERT. Мы посмотрим на некоторые варианты анализа того, что происходит внутри этой модели, а также познакомимся с ее более поздними модификациями: RoBERTa, ALBERT и другими.	Видео Презентация Основные статьи: BERT: https://arxiv.org/pdf/1810.04805.pdf BERTology: https://arxiv.org/pdf/2002.12327.pdf Analysis of BERT: https://arxiv.org/pdf/1909.00512.pdf https://arxiv.org/pdf/1908.08593v2.pdf https://arxiv.org/pdf/1906.04341v1.pdf https://arxiv.org/pdf/1908.04211.pdf RoBERTa: https://arxiv.org/abs/1907.11692 ALBERT: https://arxiv.org/pdf/1909.11942.pdf

24.04.2020	Екатерина Лобачева, научный сотрудник лаборатории компании Самсунг-НИУ ВШЭ BERT: model, analysis and modifications (Часть 2)	Видео Презентация
01.05.2020	Виктор Оганесян, младший научный сотрудник лаборатории компании Самсунг-НИУ ВШЭ Neural Stochastic Differential Equations part 2 Ввиду того, что в моделях Neural SDE и Neural ODE вводятся непрерывные по времени структуры (дифференциальные уравнения), в последнее время выходят работы, которые применяют эти модели к временным рядам. На данном семинаре мы наиболее подробно остановимся на работе https://arxiv.org/pdf/2001.01328.pdf. В ней создан аналог аджоинт метода для Neural SDE, который сильно экономит память при обучении. Также рассмотрим какие задачи пытаются решать с помощью данной модели.	Видео Презентация Статья 1, 2, 3
08.05.2020	Андрей Атанов, младший научный сотрудник лаборатории компании Самсунг-НИУ ВШЭ Contrastive Self-Supervised Learning for Image Representations На одном из предыдущих докладов мы разбирали метод обучения представлений слов для задач NLP без разметки. В данном докладе мы поговорим о self-supervised техниках предобучения для картинок. Хотя такие методы существуют давно, все они работали хуже чем предобучение на полностью размеченном датасете Iamge Net. За последний год было предложено несколько методов основанных на разных вариациях contrastive loss'a, которые работают также или лучше чем предобучение с разметкой. Мы подробно остановимся на двух методах: CPC (Contrasitve Prediction Codin g) и SimCLR (Simple Framework for Contrastive Learning).	Видео Презентация CPC: https://arxiv.org/pdf/1905.09272.pdf https://arxiv.org/pdf/1807.03748.pdf SimCLR: https://arxiv.org/pdf/2002.05709.pdf Полезные ресурсы: - https://lilianweng.github.io/lil-log/2019/11/10/self-supervised-learning.html - https://github.com/jason718/awesome-self-supervised-learning Еще модели: MoCo: https://arxiv.org/pdf/1911.05722.pdf PIRL: https://arxiv.org/pdf/1912.01991.pdf
15.05.2020	Виктор Януш, стажёр-исследователь лаборатории компании Самсунг-НИУ ВШЭ Training of binary neural networks	Видео Презентация

	На одном из предыдущих докладов обсуждалась квантизация сетей — снижение точности представлений весов и/или активаций. В данном докладе будут обсуждаться бинарные сети, в которых веса и активации могут иметь лишь два значения — +1 и - 1. Также будут разобраны различные методы обучения стохастических бинарных сетей, в которых веса и активации являются случайными величинами. Мы подробно рассмотрим новый метод, обобщающий популярные подходы и дающий им теоретическое обоснование. Также будут рассмотрены применения этого метода к обучению байесовских бинарных сетей.	Литература: 1) https://arxiv.org/abs/1603.05279 2) https://arxiv.org/abs/1602.02830 3) https://arxiv.org/abs/1812.01965
29.05.2020	Александр Лыжов, Samsung AI Center Moscow, Research Scientist Planning in Deep Reinforcement Learning Model-based deep reinforcement learning (RL) is seeing renewed interest because of promises of sample-efficiency (using less environment interactions for learning) and transferability (environment model could be reused for different tasks). The word "planning", understood broadly, refers to ways of using an environment model to improve agent training. In this talk I will 1) cover some classical planning theory and talk about pros and cons of different ways of planning, 2) talk about planning in deep RL and deconstruct MuZero - one of the most sample-efficient and highest-performing approaches to planning in complex environments with discrete timesteps and actions, 3) discuss related work and compare MuZero with competing approaches.	Презентация Sutton-Barto RL textbook chapter 8 http://incompleteideas.net/book/RLbook2020.pdf MuZero https://arxiv.org/abs/1911.08265 Видео

Осенний семестр 2019 г.

Дата	Докладчик и тема	Материалы
13.09.2019	Артём Соболев, Samsung AI Center Moscow, Research Scientist On Mutual Information Estimation Mutual Information is an important information-theoretic concept that captures an intuitive idea of the amount of information shared between two random variables. Mutual Information has been used extensively in numerous Machine Learning problems and should be of great interest for every ML researcher. In practice, however, accurately estimating the Mutual Information is a non-trivial task. Recently, it has been shown that many general estimators fail to produce reasonable estimates unless an exponential number of samples is taken. We will discuss this result with its manifestation in several widely used estimators, and then consider new estimators that sidestep the core issue.	Презентация; Видео
20.09.2019	Сергей Трошин, НИУ ВШЭ Deep Equilibrium Models	Презентация; Статья 1, 2, 3, 4;

	Very deep neural networks can require a lot of memory to be stored for the backpropagation. We will consider a recently proposed approach for modelling sequential data: the deep equilibrium model (DEQ). It can be observed that for some deep models layers' outputs tend to converge to a fixed point with the increase of the network's depth. The DEQ approach directly finds these equilibrium points via root-finding. We will see how to analytically backpropagate through the equilibrium point using implicit differentiation which proves to be very memory efficient.	Видео
27.09.2019	Максим Кодрян, стажер-исследователь лаборатории компании Самсунг-НИУ ВШЭ Invariant Risk Minimization This talk is dedicated to the correlation-versus-causation dilemma. Minimizing training error leads machines into recklessly absorbing all the correlations found in training data. Understanding which patterns are actually useful (causal) is important if we want our models to generalize to new test distributions. It seems that there exists an intimate link between invariance and causation useful for generalization. We will consider the concept of Invariant Risk Minimization (as opposed to Empirical Risk Minimization) — a novel learning paradigm that estimates nonlinear, invariant, causal predictors from multiple training environments, to enable out-of-distribution generalization. We will also provide an information-theoretic view on the topic.	Презентация; Статья ; Блог-пост; Видео
04.10.2019	Евгений Голиков, исследователь лаборатории нейронных систем и глубокого обучения МФТИ Why do neural nets learn and generalize? As was noted in [Belkin et al., 2019], neural nets are usually used in the so-called "interpolating regime". In this regime our architecture is large enough to have an ability to fit the training data perfectly, as opposed to "classical regime", where our model is constrained to balance between learning and generalization. Two questions arise immediately: 1) Why does (stochastic) gradient descent - a local optimization method - find a configuration that fits the data perfectly? 2) Why does (stochastic) gradient descent choose a configuration that generalize well, across all configurations that fit the training data? Although the first question is close to being fully answered, the second one remains mostly opened. In our talk we will review some of the recent results concerning both of them.	Презентация; Статья 1, Статья 2, Статья 3, Статья 4, Статья 5; Видео
11.10.2019	Евгений Егоров, студент Сколтеха The Implicit Metropolis-Hastings Algorithm Recent works propose using the discriminator of a GAN to filter out unrealistic samples of the generator. We generalize these ideas by introducing the implicit Metropolis-Hastings algorithm. For any implicit probabilistic model and a target distribution represented by a set of samples, implicit Metropolis-Hastings operates by learning a discriminator to estimate the density-ratio and then generating a chain of samples. Since the approximation of density ratio introduces an error on every step of the chain, it is crucial to analyze the stationary distribution of such chain. For that purpose, we present a theoretical result stating that the discriminator loss upper bounds the total variation distance between the target distribution and the stationary distribution. Finally, we validate the proposed algorithm both for independent and Markov proposals on CIFAR-10 and CelebA datasets.	Презентация; Статья

18.10.2019	Андрей Малинин, старший исследователь, Yandex Research Reverse KL-Divergence training of Prior Networks Ensemble approaches for uncertainty estimation have recently been applied to the tasks of misclassification detection, out-of-distribution input detection and adversarial attack detection. Prior Networks have been proposed as an approach to efficiently emulate an ensemble of models for classification by parameterising a Dirichlet prior distribution over output distributions. These models have been shown to outperform alternative ensemble approaches, such as Monte-Carlo Dropout, on the task of out-of-distribution input detection. However, scaling Prior Networks to complex datasets with many classes is difficult using the training criteria originally proposed. This paper makes two contributions. First, we show that the appropriate training criterion for Prior Networks is the reverse KL-divergence between Dirichlet distributions. This addresses issues in the nature of the training data target distributions, enabling prior networks to be successfully trained on classification tasks with 200 classes, as well as improving out-of-distribution detection performance. Second, taking advantage of this new training criterion, this paper investigates using Prior Networks to detect adversarial attacks. It is shown that the construction of successful adaptive whitebox attacks, which affect the prediction and evade detection, against Prior Networks trained on CIFAR-10 and CIFAR-100 takes a greater amount of computational effort than against standard neural networks, adversarially trained neural networks and dropout-defended networks.	Презентация; Статья о Prior Networks; Статья (будет представлена на NeurIPS) ; Видео
25.10.2019	Андрей Леонидов (ФИАН, МФТИ) Фазовые переходы в байесовском оценивании В докладе обсуждается анализ фазовых переходов в байесовском оценивании с использованием методов статистической физики. Основное внимание уделяется анализу фазового перехода легкое-трудное оценивание (easy-hard inference), сопровождающегося возникновением фазы стекла (glass phase).	
01.11.2019	Айбек Аланов, Samsung AI Center Moscow, Research Engineer Implicit λ-Jeffreys Autoencoders: Taking the Best of Both Worlds We propose a new form of an autoencoding model which incorporates the best properties of variational autoencoders (VAE) and generative adversarial networks (GAN). It is known that GAN can produce very realistic samples while VAE does not suffer from mode collapsing problem. Our model optimizes λ-Jeffreys divergence between the model distribution and the true data distribution. We show that it takes the best properties of VAE and GAN objectives. It consists of two parts. One of these parts can be optimized by using the standard adversarial training, and the second one is the very objective of the VAE model. However, the straightforward way of substituting the VAE loss does not work well if we use an explicit likelihood such as Gaussian or Laplace which have limited flexibility in high dimensions and are unnatural for modelling images in the space of pixels. To tackle this problem we propose a novel approach to train the VAE model with an implicit likelihood by an adversarially trained discriminator. In an extensive set of experiments on CIFAR-10 and TinyImagent datasets, we show that our model achieves the state-of-the-art trade-off between generation and reconstruction quality and demonstrate how we can balance between mode-seeking and mass-covering behaviour of our model by adjusting the weight λ in our objective.	Видео; Статья; Презентация
08.11.2019	Не проводим семинар	
15.11.2019	Арсений Ашуха, PhD Candidate at Bayesian Methods Research Group & Samsung AI	Статья 1;

	The state of uncertainty Uncertainty estimation and ensembling methods go hand-in-hand. Uncertainty estimation is one of the main benchmarks for assessment of ensembling performance. At the same time, deep learning ensembles have provided state-of-the-art results in uncertainty estimation. In the talk, we will consider the most popular metrics for in-domain uncertainty estimates and its pitfalls and fixes. We will discuss the results of a broad study of different ensembling techniques, and introduce the deep ensemble equivalent---a new metric that allows us to compare the result of ensembling between different architectures and datasets. We will see that many sophisticated ensembling techniques are equivalent to an ensemble of very few independently trained networks. Depending on available time we will cover the study of out-of-domain uncertainty by Ovadia2019.	Презентация; Видео
22.11.2019	Александр Фрей, Researcher at NORMENT (Norwegian Centre for Mental Disorders Research), University of Oslo, Norway Mathematical models of the genetic architecture in complex human disorders Modern studies on genetics of complex human disorders collect large samples, often exceeding N=10^6 individuals and M=10^7 genetic variants, posing challenging mathematical problems, such as solving a system of linear equations with huge NxM design matrix. In this presentation we will describe the Gaussian Mixture model (MiXeR [1], [2]) and three approaches for estimating its probability density function using (1) random sampling, (2) Fourier convolution, and (3) moment-preserving approximations. Further, we discuss our optimization protocol, based on direct maximization of the likelihood function using differential evolution and Nelder-Mead algorithms. Finally, we derive posterior estimates for some quantities of interest. If time allows we may also discuss related work [3] based on Mixed Linear Models, REML (Restricted Maximum Likelihood) and Variational iteration for Bayesian linear regression with Gaussian mixture prior.	Презентация; Видео; Литература: [1] AI-MiXeR (most important details are in the Online Methods section) [2] Cross-trait MiXeR (most important details are in the supplementary note here) [3] Efficient Bayesian mixed-model analysis increases association power in large cohorts (Nature Genetics, 2015)
29.11.2019	Екатерина Шульман	Видео
06.12.2019	Не проводим семинар	
13.12.2019	Не проводим семинар	
20.12.2019	Полина Кириченко и Павел Измаилов, New York University Scalable Bayesian inference in low-dimensional subspaces Bayesian methods can provide full-predictive distributions and well-calibrated uncertainties in modern deep learning. However, scaling Bayesian inference techniques to deep neural networks (DNNs) is challenging due to the high	Презентация; Статья 1; Статья 2; Видео

	dimensionality of the parameter space. In this talk, we will discuss two recent papers on scalable Bayesian inference which share a similar high-level idea: performing approximate inference in low-dimensional subspaces of DNNs parameter space. In Subspace Inference for Bayesian Deep Learning [1], we propose to exploit the geometry of DNN training objectives to construct low-dimensional subspaces that contain diverse sets of models. In these subspaces, we are able to apply a wide range of advanced approximate inference methods, such as elliptical slice sampling and variational inference, that struggle in the full parameter space. We show that Bayesian model averaging over the induced posterior in these subspaces leads to strong performance in terms of accuracy and uncertainty quantification on regression and image classification tasks. In Projected BNNs [2], the authors propose a variational inference framework for Bayesian neural networks that (1) encodes complex distributions in high-dimensional parameter space with representations in a low-dimensional latent space, and (2) performs inference efficiently on the low-dimensional representations.	

Весенний семестр 2019 г.

Дата	**Докладчик и тема**	Материалы
15.02.2019	Тимур Гарипов, Samsung AI Center Moscow, Engineer SWAG: Approximate Bayesian Inference Using SGD Trajectory We propose SWA-Gaussian (SWAG), a simple, scalable, and general purpose approach for uncertainty representation and calibration in deep learning. Stochastic Weight Averaging (SWA), which computes the first moment of stochastic gradient descent (SGD) iterates with a modified learning rate schedule, has recently been shown to improve generalization in deep learning. With SWAG, we fit a Gaussian using the SWA solution as the first moment and a low rank plus diagonal covariance also derived from the SGD iterates, forming an approximate posterior distribution over neural network weights; we then sample from this Gaussian distribution to perform Bayesian model averaging. We empirically find that SWAG approximates the shape of the true posterior, in accordance with results describing the stationary distribution of SGD iterates. Moreover, we demonstrate that SWAG performs well on a wide variety of computer vision tasks, including out of sample detection, calibration, and transfer learning, in comparison to many popular alternatives including MC dropout, KFAC Laplace, and temperature scaling.	Презентация; Статья; Видео
22.02.2019	Виктор Оганесян, МФТИ, Институт высшей нервной деятельности Neural Ordinary Differential Equations This talk is based on the first part of the paper "Neural ordinary differential equations". Authors introduce a concept of residual networks with continuous-depth, what they consider as ordinary differential equations (ODEs). Correspondingly, inputs of neural networks are considered as an initial state of ODEs, and outputs as a solution obtained by ODE solver. One	Презентация; Статья; Видео

	of the main advantages of such approach is the constant memory cost with respect to the model depth. However, training of such networks requires introduction of adjoint function (standard technique from optimal control theory). One of the curious points is that solving of ODEs for the adjoint function can be considered as continuous analog of backpropagation.	
01.03.2019	Александра Волохова, МФТИ, ШАД Continuous Normalizing Flows My presentation will be a continuation of Victor's talk about Neural ODE. I'll explain how this idea can be applied to normalizing flows, making them more flexible and computable. Furthermore, we will talk about FFJORD — an unbiased stochastic estimator of the likelihood based on continuous normalizing flows. As authors of the paper state, this approach allows creating reversible generative models with completely unrestricted architectures.	Презентация; Статья 1; Статья 2; Видео
15.03.2019	Максим Кузнецов, Сколтех, Insilico Medicine A Tensor Ring Induced Prior for Generative Models Generative models produce realistic objects in many domains, including text, image, video, and audio synthesis. Most popular models — Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) — usually employ a standard Gaussian distribution as a prior. Previous works show that richer family of prior distributions may help to avoid the mode collapse problem in GANs and to improve the evidence lower bound in VAEs. We propose a new family of prior distributions — Tensor Ring Induced Prior (TRIP) — that packs an exponential number of Gaussians into a high-dimensional lattice with a relatively small number of parameters. We show that these priors improve Frechet Inception Distance for GANs and Evidence Lower Bound for VAEs. We also study generative models with TRIP in the conditional generation setup with missing conditions. Altogether, we propose a novel plug-and-play framework for generative models that can be utilized in any GAN and VAE-like architectures.	
22.03.2019	Артём Соболев, Samsung AI Center Moscow, Research Scientist Importance Weighted Hierarchical Variational Inference Variational Inference is a powerful tool in the Bayesian modeling toolkit, however, its effectiveness is determined by the expressivity of the utilized variational distributions in terms of their ability to match the true posterior distribution. Recently, there's been a lot of work on employing neural networks as powerful sample generators, but the need for a tractable density is a major limitation. In talk I will suggest a novel method based on a multisample variational bound that generalizes many previous works, most importantly, Hierarchical Variational Models and Semi-Implicit Variational Inference. The bound allows us learn more expressive approximate posteriors, and can be combined with many prior results.	Презентация; Статья; Видео
29.03.2019	Виктор Руднев, стажер-исследователь лаборатории компании Самсунг-НИУ ВШЭ, МГУ им. М.В. Ломоносова Sparse Bayesian Variational Learning with Matrix Normal Distributions The application of variational Bayesian methods to neural networks has been limited by the choice of the posterior approximation family. One could use a simple family like a normal distribution with independent variables, but that results in a low quality of the approximation and optimization issues. In the paper we propose to use Matrix Normal distribution (MN) for variational approximation family. While being more flexible, this family supports efficient reparameterization and Riemannian	Презентация; Статья; Видео

	optimization procedures. We apply this family for Bayesian neural networks sparsification through Automatic Relevance Determination (Kharitonov et al., 2018). We show that MN family here outperforms simpler fully-factorized Gaussians, especially for the case of group sparsification, while remaining as computationally efficient as the latter. We also analyze application of MN distribution for inference in Variational Auto-Encoder model.	
05.04.2019	Ануар Таскынов, МГУ им. М.В. Ломоносова Introduction to riemannian optimization and its application on matrix manifolds Riemannian optimization is a new point of view is offered for the solution of constrained optimization problems. Some classical optimization techniques on Euclidean space are generalized to Riemannian manifolds. This talk consist of two parts: 1) basic notions of differential geometry, 2) optimization algorithms on matrix manifolds.	Презентация; Литература (Optimization Algorithms on Matrix Manifolds); Видео
12.04.2019	Андрей Леонидов, доктор физ.-мат. наук, ФИАН, МФТИ Спиновые стёкла - основы теории В докладе обсуждаются основы современного теоретического описания свойств спиновых стекол (магнетиков со случайными взаимодействиями). Важное значение теории спиновых стекол для задач computer science связано с тем, что статистическая физика спиновых стекол является универсальной метафорой для сложных оптимизационных задач, в которых целевая функция имеет большое количество близких по величине максимумов.	Видео
19.04.2019	Даниил Полыковский, Insilico Medicine The Kanerva Machine The talk will cover the recently proposed Kanerva Machine—a model that employs associative memory, in contrast to a slot-based memory. Kanerva Machine views memory as a random variable and can make Bayesian inference of its content. With a chosen parameterization of memory, authors could do fast iterative writing using the Bayes formula. In the talk, we will also see further development of this model in learning attractor dynamics, useful for applications like denoising.	Презентация; Статья 1; Статья 2; Видео
26.04.2019	Арсений Кузнецов, Samsung AI Center Moscow, Engineer Reinforcement learning for POMDP via Variational inference and Particle filtering Control in a Partially Observable Markov Decision Processes (POMDPs) relies on the sequence of observations that carry only partial information about underlying Markov state. One way to integrate the observation sequence into fixed size representation is Bayesian filtering. This talk will cover variational sequential filtering and it's application to reinforcement learning in POMDP via maximising the variational lower bound on the log marginal likelihood of observations, rewards, and surrogate optimality variables.	Презентация; Статья 1; Статья 2; Статья 3; Статья 4

17.05.2019	Павел Темирчев, PhD Student, Research Intern; Skoltech, Centre for Hydrocarbon Recovery Reinforcement Learning as Probabilistic Inference For the past few years, RL has shown huge progress in solving simulated tasks, such as Go, Dota, Atari and Starcraft. Though RL still is not broadly applicable for physical agents, such as robots, due to the huge sample complexity and local optimality of learned policies. One way to improve standard RL is to enforce policies to be as random as possible. We will show that it might be achieved via a probabilistic look on the problem. Inference in Markov process augmented with optimality variables can be shown to be equivalent to the so-called Maximum Entropy RL framework. We will look on a variational inference procedure within our graphical model and will derive soft analogues of Q and V value functions (known from the standard RL approach), which can be used in soft versions of Q-learning and Actor-Critic algorithms able to produce diverse and multimodal policies. The probabilistic framework opens doors for very new algorithms and ideas such as hierarchical policies, which are also will be discussed.	Презентация; Статья 1; Статья 2; Статья 3; Статья 4

Осенний семестр 2018 г.

Дата	Докладчик и тема	Материалы
14.09.2018	Дмитрий Молчанов, AIC Samsung Research, Senior Engineer; научный сотрудник Лаборатории компании Самсунг-НИУ ВШЭ Variational inference with implicit distributions Conventional variational inference problems are defined by the likelihood function, the prior distribution and the parametric approximate posterior, which are all usually explicit: we can sample from them, reparameterize them and compute their density. As soon as one component becomes implicit (we can't compute the density), the variational inference becomes intractable. In this talk I will review several approaches that allow us to perform variational inference with implicit distributions. The use of implicit variational inference provides many exciting benefits from fitting an arbitrarily flexible implicit posterior to likelihood-free variational inference.	Презентация; Литература: arxiv1; arxiv2; arxiv3; Видео
21.09.2018	Павел Швечиков, AIC Samsung Research, Senior Engineer Variational Sequential Monte Carlo Many reliable algorithms exist for sequential Bayesian inference in simple settings, such as time series with discrete latent states (tackled by HMM) or models with latent linear-Gaussian dynamics (tackled by Kalman filter). In practice, however, the	Презентация; Литература: arxiv1; arxiv2; arxiv3;

	sequences we most often face have complex, nonlinear, non-Gaussian dependencies and are high-dimensional. How to perform an accurate inference in such a setting? The question has been investigated for about thirty years and has given rise to a variety of sophisticated approximate approaches (such as Extended Kalman Filter and Gaussian-sum filter). The true breakthrough in the field was made by introducing the concept of Particle Filter (PF) in the mid-90s. The concept underlying the PF – Sequential Monte Carlo (SMC) – made it possible to significantly extend the scope of "solvable" tasks in computer vision, financial econometrics, target tracking, robotics, geosciences, system biology, and many other fields. In the talk, we will get acquainted with the basics of SMC and learn about its recent applications for deep generative modeling that get us closer to fast, scalable and accurate Bayesian inference in both sequential and non-sequential settings.	Видео
28.09.2018	Даниил Полыковский, Insilico Medicine Deep Learning for Drug Discovery Neural networks and other machine learning models have recently been applied to many biological problems, including drug discovery. In this field, different kinds of generative models were applied to generate novel molecular structures in forms of strings and graphs. Along with the general toolbox of neural networks, multiple novel ideas were introduced to build generators of molecules, including models working with data represented as graphs. In my talk, I will give an overview of the drug discovery pipeline and how machine learning can be applied on each step. I will also cover many novel ideas and tricks used in this field, that can be extended to other domains.	Презентация; Литература (по темам): Optimization; Conditional 1, 2; Reinforcement Learning 1, 2, 3; Strings 1, 2; Graphs 1, 2, 3; 3D; Видео
05.10.2018	Кирилл Неклюдов, Machine Learning Engineer, AIC Samsung Research; научный сотрудник Лаборатории компании Самсунг-НИУ ВШЭ Optimization of proposal distribution for the Metropolis-Hastings algorithm In this paper we propose to view the acceptance rate of the Metropolis-Hastings algorithm as a universal objective for learning to sample from target distribution -- given either as a set of samples or in the form of unnormalized density. This point of view unifies the goals of such approaches as Markov Chain Monte Carlo (MCMC), Generative Adversarial Networks (GANs), variational inference. To reveal the connection we derive the lower bound on the acceptance rate and treat it as the objective for learning explicit and implicit samplers. The form of the lower bound allows for doubly stochastic gradient optimization in case the target distribution factorizes (i.e. over data points). We empirically validate our approach on Bayesian inference for neural networks and generative models for images. TL;DR: Learning to sample via lower bounding the acceptance rate of the Metropolis-Hastings algorithm	Презентация; Статья; Видео
12.10.2018	Андрей Атанов, стажер-исследователь Лаборатории компании Самсунг-НИУ ВШЭ, магистр ФКН НИУ ВШЭ Deep Weight Prior Bayesian inference is known to provide a general framework for incorporating prior knowledge or specific properties into machine learning models via carefully choosing a prior distribution. In this work, we propose a new type of prior distributions for convolutional neural networks, deep weight prior, that in contrast to previously published techniques, favors empirically estimated structure of convolutional filters e.g., spatial correlations of weights. We define deep weight prior as an implicit distribution and propose a method for variational inference with such type of implicit priors. In experiments, we show that deep weight priors can improve the performance of Bayesian neural networks on several problems when training data is limited. Also, we found that initialization of weights of conventional convolutional networks with samples from deep weight	Презентация; Статья; Видео

	prior leads to faster training	
19.10.2018	Валерий Харитонов, младший научный сотрудник Лаборатории компании Самсунг-НИУ ВШЭ, аспирант ФКН НИУ ВШЭ (Doubly) Semi-Implicit Variational Inference This is a follow-up to the first talk of this year in which I will tell you more about variational inference with implicit distributions. This time, we will assume that the approximate posterior and the prior can be both expressed as an intractable infinite mixture of some analytic density with a highly flexible implicit mixing distribution. It turns out, this formulation allows one to perform both variational inference and variational learning and gives a sandwich bound on the ELBO which is asymptotically exact. At the end of the talk, I will tell you a bit about the use cases for (doubly) semi-implicit variational inference and learning and our experimental results.	Презентация; Статья 1; Статья 2; Видео
26.10.2018	Кирилл Струминский, аспирант ФКН НИУ ВШЭ, стажер-исследователь Центра глубинного обучения и байесовских методов НИУ ВШЭ Quantifying Learning Guarantees for Convex but Inconsistent Surrogates, to appear at NIPS 2018 We study consistency properties of machine learning methods based on minimizing convex surrogates. We extend the recent framework of Osokin et al. [Статья 1] for quantitative analysis of the consistency properties to the case of inconsistent surrogates. Our key technical contribution consists in the new lower bound on the calibration function for the quadratic surrogate, which is non-trivial (not always zero) for inconsistent cases. The new bound allows to quantify the level of inconsistency of the setting and shows how learning with inconsistent surrogates can have guarantees on sample complexity and optimization difficulty. We apply our theory in two concrete cases: multi-class classification with the tree-structured loss and ranking with the mean average precision loss. The results show the approximation-computation trade-offs caused by inconsistent surrogates and their potential benefits.	Презентация; Статья 1; Видео
02.11.2018	Виктор Януш, стажер-исследователь Лаборатории компании Самсунг-НИУ ВШЭ Hamiltonian Monte-Carlo for Orthogonal Matrices We consider the problem of sampling from posterior distributions for Bayesian models where some parameters are restricted to be orthogonal matrices. Such matrices are sometimes used in neural networks models for reasons of regularization and stabilization of training procedures, and also can parameterize matrices of bounded rank, positive-definite matrices and others. We propose a new sampling scheme that is based on Hamiltonian Monte Carlo (HMC) approach and ideas of Riemannian optimization for a set of orthogonal matrices. The method is theoretically justified by proof of symplecticity for the proposed iteration. In experiments we show that the new scheme is more sample-efficient comparing to conventional HMC with explicit orthogonal parameterization. We also provide promising results of Bayesian ensembling for orthogonal neural networks and low-rank matrix factorization.	Презентация; Статья 1, 2, 3; Видео
09.11.2018	Лекция про Первую Мировую войну :-)	Видео
16.11.2018	Павел Темирчев, PhD Student, Research Intern; Skoltech, Centre for Hydrocarbon Recovery Predicting Oil Movement in a Development System using Deep Latent Dynamics Models. We present a novel technique for assessing the dynamics of multiphase fluid flow in the oil reservoir. We demonstrate an efficient workflow for handling the 3D reservoir simulation data in a way which is orders of magnitude faster than the	Презентация; Статья 1, 2;

	conventional routine. The workflow (we call it "Metamodel") is based on projecting the dynamical system into nonlinear subspace where the dynamics is captured by deep recurrent neural network. Compared to basic reduced order modelling approaches our projecting technique involves usage of variational autoencoder model instead of linear ones. We show that being trained on multiple results of the conventional reservoir modelling, the Metamodel does not compromise the accuracy of the reservoir dynamics reconstruction in a significant way. It allows forecasting not only the flow rates from the wells but also the dynamics of the distribution of pressure and fluid saturations within the reservoir. The results open a new perspective in the optimization of oilfield development as the scenario screening could be accelerated sufficiently. During the talk, I will introduce you to the classical POD-Galerkin approach to reduce the computational cost of modelling multi-phase flows through a porous medium, to the recently published reduced order model based on POD and Deep Residual RNNs and to the Metamodelling technique proposed by us. No background on oil field development routine is needed - I will make a small intro to the task.	Видео
23.11.2018	Diego Granziol, researcher, Oxford-Man Institute of Quantitative Finance Title 1:Maximum Entropy and learning the spectra of massive graphs. Abstract 1: The method of maximum entropy with its origin in statistical mechanics and information theory has found many uses in machine learning and has formed a natural prior for Bayesians. We consider the convergence of the moments the spectral density to the underlying stochastic process using the machinery of random matrix theory. We show that the method of maximum entropy forms a natural basis for measuring the divergence between graphs and that kernel smoothing techniques are information destroying. Title 2: Learning the Spectra of Deep Neural networks with application to learning the learning rate and the momentum Abstract 2: We develop a methodology that allows us to analyse and visualise the loss surface of networks with millions or tens of millions of parameters. By considering the network spectrum evolution and the convergence of GD with Momentum on a convex quadratic, along with some random matrix theory, we introduce a method for learning the learning rate and momentum in deep nets.	Презентация; Видео
30.11.2018	Айбек Аланов, AIC Samsung Research, Machine Learning Engineer Pairwise Augmented GANs with Adversarial Reconstruction Loss Аннотация: We consider a problem of training bidirectional GANs. We propose a novel autoencoding model called Pairwise Augmented GANs. We train a generator and an encoder jointly and in an adversarial manner. The generator network learns to sample realistic objects. In turn, the encoder network at the same time is trained to map the true data distribution to the prior in latent space. To ensure good reconstructions, we introduce an augmented adversarial reconstruction loss. Here we train a discriminator to distinguish two types of pairs: an object with its augmentation and the one with its reconstruction. We show that such adversarial loss compares objects based on the content rather than on the exact match. We experimentally demonstrate that our model generates samples and reconstructions of quality competitive with state-of-the-art on datasets MNIST, CIFAR10, CelebA and achieves good quantitative results on CIFAR10.	Презентация; Статья; Видео
07.12.2018	NIPS	
14.12.2018	Александр Шевченко, стажер-исследователь Лаборатории компании Самсунг-НИУ ВШЭ Scaling Matters in Deep Structured-Prediction Models Deep structured-prediction energy-based models combine the expressive power of learned representations and the possibility of embedding knowledge about the task at hand into the system. A common way to learn parameters of such	Презентация; Видео

	models consists in a multistage procedure where different combinations of components are trained at different stages. The joint end-to-end training of the whole system is then done as the last fine-tuning stage. This multistage approach is time-consuming and cumbersome as it requires multiple runs until convergence and multiple rounds of hyperparameter tuning. From this point of view, it is beneficial to start the joint training procedure from the beginning, however, such approaches often unexpectedly fail and deliver results worse than the multistage ones. In this paper, we hypothesize that one reason for joint training of deep energy-based models to fail consists in the incorrect relative normalization of different components in the energy function. We propose online and offline scaling algorithms that fix the joint training and demonstrate their efficacy on three different tasks.	
21.12.2018	зачет	
28.12.2018		

Весенний семестр 2018 г.

Дата	**Докладчик и тема**	Материалы
16.02.2018	Разбор статей ICLR-2018 Александр Новиков, аспирант ИВМ РАН Distributional Policy Gradients Антон Родоманов, аспирант ФКН НИУ ВШЭ On the Convergence of Adam and Beyond Юрий Кемаев, магистр ФКН НИУ ВШЭ, Сколтех Compressing Word Embeddings via Deep Compositional Code Learning	слайды_Новиков, слайды_Родоманов, слайды_Кемаев
02.03.2018	Андрей Атанов, студент ФКН НИУ ВШЭ Stochastic Batch Normalization In this work, we investigate Batch Normalization technique and propose its probabilistic interpretation. We propose a probabilistic model and show that Batch Normalization maximazes the lower bound of its marginalized log-likelihood. Then, according to the new probabilistic model, we design an algorithm which acts consistently during train and test. However, inference becomes computationally inefficient. To reduce memory and computational cost, we propose Stochastic Batch Normalization -- an efficient approximation of proper inference procedure. This method provides us with a scalable uncertainty estimation technique.	презентация видео
16.03.2018	Aliaksandr Hubin (University of Oslo) Deep Bayesian regression models Regression models are addressed for inference and prediction in a wide range of applications providing a powerful scientific tool for the researchers and analysts coming from different fields. In most of these fields more and more sources of data are becoming available introducing a variety of hypothetical explanatory variables for these models to be considered. Model averaging induced by	презентация видео

	different combinations of these variables becomes extremely important for both good inference and prediction. Not less important, however, seems to be the quality of the set of explanatory variables to select from. It is often the case that linear relations between the explanatory variables and the response are not sufficient for the high quality inference or predictions. Introducing non-linearities and complex functional interactions based on the original explanatory variables can often significantly improve both predictive and inferential performance of the models. The non-linearities can be handled by deep learning models. These models, however, are often very difficult to specify and tune. Additionally they can often experience over-fitting issues. Random effects are also not incorporated in the existing deep learning approaches. In this paper we introduce a class of deep Bayesian regression models with latent Gaussian variables generalizing the classes of GLM, GLMM, ANN, CART, logic regressions and fractional polynomials into a powerful and flexible Bayesian framework. We then suggest algorithmic approaches for fitting them. In the experimental section we test some computational properties of the algorithm and show how deep Bayesian regression models can be used for inference and predictions in various applications.	
23.03.2018	Тимур Гарипов, ВМК МГУ, ФКН НИУ ВШЭ Loss Surfaces, fast ensembling and weight averaging of DNNs The loss functions of DNNs are complex and their geometric properties are not well understood. We show that the optima of these complex loss functions are in fact connected by a simple curve over which training and test accuracy are nearly constant. We introduce a training procedure to discover these high-accuracy pathways between modes. Inspired by this new geometric insight, we propose a new ensembling method entitled Fast Geometric Ensembling (FGE). Using FGE we can train high-performing ensembles in the time required to train a single model. Visualizing loss surfaces containing networks from FGE ensembles we noticed that the simple averaging of points from these ensembles leads to better generalization than conventional training. We proposed a method of neural network training that is called SWA and leads to wider local optima. Using SWA we achieve notable improvement in test accuracy over conventional SGD training on a range of state-of-the-art residual networks, PyramidNets, DenseNets, and Shake-Shake networks on CIFAR-10, CIFAR-100, and ImageNet.	презентация; arxiv; arxiv2 видео
30.03.2018	Евгений Никишин, магистр ФКН НИУ ВШЭ, Сколтех Hierarchical methods for Reinforcement Learning Hierarchical Reinforcement Learning aims to operate in terms of macro-actions or high-level goals. Motivation behind this is the following: if we are given meaningful goals or macro-actions, it will accelerate learning, increase interpretability of policy and grant faster exploration. Originally, it was proposed to design macro-actions or goals manually, so the most challenging part of today's HRL research is learning options purely from interaction with environment. In this talk, we will discuss classic approaches for HRL, when some elements of policy were provided by human, as well as recently proposed methods for HRL without any prior knowledge of environment.	презентация; arxiv; arxiv2 видео
06.04.2018	Павел Швечиков, аспирант ФКН НИУ ВШЭ Learning in Partially Observable Markov Decision Processes Probably, you have already been tired of news headlines of the kind "Artificial Intelligence has defeated humankind!", actually implying that some specific algorithm has been tuned to beat the top human performer in a game X. Such algorithms, though interesting by themselves, are not viable in a real life (for example, see the post https://medium.com/@karpathy/alphago-in-context-c47718cb95a5 for analyzing AlphaGo's applicability outside of Go). The real world is far more complex than the games with full information. At the very least, the world has the fundamental property of being Partially Observable. Thus it is essential for an intelligent agent of Future to be able to act optimally under very restrictive conditions of concealed information. In this talk, I will explain why learning in partially observable environments is so hard and what are some of the approaches to deal with it.	презентация видео

13.04.2018	Арсений Ашуха Multi-agent reinforcement learning Reinforcement Learning has become wide and important topic of machine learning research. Despite great success was achieved in single-agent environments, multi-agent problem setting is not studied well. Multiagent environments require not just ability to learn several agents simultaneously, but also to communicate and cooperate in order to find individual strategies beyond out of selfishness in order to achieve a high joined reward. In Fridays talk we will address two topics: a) Sequential Social Dilemmas that allow you to study cooperation [1] and to learn agents to cooperate [2] in multiagent-environments b) Nonstationarity issues, namely, how to avoid nonstationarity in multiagent environments [3, 4]. [1] Multi-agent Reinforcement Learning in Sequential Social Dilemmas https://arxiv.org/abs/1702.03037 [2] Inequity aversion resolves intertemporal social dilemmas https://arxiv.org/abs/1803.08884 [3] Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments https://arxiv.org/abs/1706.02275 [4] Stabilising Experience Replay for Deep Multi-Agent Reinforcement Learning https://arxiv.org/abs/1702.08887	презентация видео
20.04.2018	Александр Горский, ИППИ РАН New critical and collective phenomena in random networks and graphs I will discuss new critical and collective phenomena found recently for the different ensembles of exponential networks. In particular the network clusterization transition induced by 3-cycles as well as the phase transitions induced by the triads which amount to synchronization and bipartiteness will be explained for the multilayer networks. The phenomena of the spontaneous symmetry breaking in the multilayer networks will be demonstrated. I shall describe the possible application of generalized Schelling model for the social dynamics.	презентация видео
27.04.2018	Репетиция выступлений на Data Fest Сбербанка	
04.05.2018	Олег Иванов, магистр ВМК МГУ Universal Conditional Machine Variational Autoencoder (VAE) is one of the most popular generative models nowadays. Despite the fact that VAE allows generating objects and compting approximate probabilistic density function for given objects, it lacks the ability to be conditioned on the arbitrary subset of the object features. We proposed the generalization of VAE called Universal Conditional Machine (UCM) to overcome this issue. In this talk we will discuss this model and its relations with other conditional generative models.	презентация видео
11.05.2018	Дмитрий Молчанов, научный сотрудник ФКН НИУ ВШЭ Variance Networks During this talk, I will introduce variance networks, a model that stores the learned information in the variances of the network weights. Surprisingly, no information gets stored in the expectations of the weights, therefore if we replace these weights with their expectations, we would obtain a random guess quality prediction. We will discuss how and why this model works, and will see how it naturally arises in several types of Bayesian Neural Networks. Then we will discuss a hueristic that uses the loss curvature to determine which random variables can be replaced with their expected values, and see that only a small fraction of weights is needed for ensembling. The success of this model raises several counter-intuitive implications for the training and application of Deep Learning models.	презентация видео
18.05.2018	Нет спецсеминара	
25.05.2018	Нет спецсеминара	

Айбиэмовцы - на осень
Полыковский - на конец осени: Deep Learning for Drug Discovery
Денис Беломестный, Сколтех (MCMC) - на осень, план

Осенний семестр 2017 г.

Дата	Докладчик и тема	Материалы
08.09.17	Антон Осокин, доцент ФКН ВШЭ On Structured Prediction Theory with Calibrated Convex Surrogate Losses We provide novel theoretical insights on structured prediction in the context of efficient convex surrogate loss minimization with consistency guarantees. For any task loss, we construct a convex surrogate that can be optimized via stochastic gradient descent and we prove tight bounds on the so-called "calibration function" relating the excess surrogate risk to the actual risk. In contrast to prior related work, we carefully monitor the effect of the exponential number of classes in the learning guarantees as well as on the optimization complexity. As an interesting consequence, we formalize the intuition that some task losses make learning harder than others, and that the classical 0-1 loss is ill-suited for structured prediction. Joint work with Francis Bach and Simon Lacoste-Julien	arxiv, слайды, видео
15.09.17	Ilya Tolstikhin, Research Scientist, Max Planck institute for Intelligent Systems AdaGAN: Boosting Generative Models Generative Adversarial Networks (GAN) are an effective method for training generative models of complex data such as natural images. However, they are notoriously hard to train and often suffer from the problem of missing modes where the model is not able to produce examples in certain regions of the space. In this talk I will present an iterative procedure, called Adaptive GAN (AdaGAN), where at every step we add a new component into a mixture model by running a GAN algorithm on a reweighted training sample. This procedure is inspired by boosting algorithms, where many potentially weak individual predictors are greedily aggregated to form a strong composite predictor. Based on the approximation bounds for general f-divergences we derive sufficient conditions for this iterative procedure to converge to the true data distribution at an exponential rate.	arxiv, слайды, видео
22.09.17	Семинара не будет	
29.09.17	Кирилл Неклюдов, аспирант ФКН ВШЭ Structured Bayesian Pruning via Log-Normal Multiplicative Noise Dropout-based regularization methods can be regarded as injecting random noise with pre-defined magnitude to different parts of the neural network during training. It was recently shown that Bayesian dropout procedure not only improves generalization but also leads to extremely sparse neural architectures by automatically setting the individual noise magnitude per weight. However, this sparsity can hardly be used for acceleration since it is unstructured. In the paper, we propose a new Bayesian model that takes into account the computational structure of neural networks and provides structured sparsity, e.g. removes neurons and/or convolutional channels in CNNs. To do this, we inject noise to the neurons outputs while keeping the weights unregularized. We established the probabilistic	arxiv, слайды, видео

	model with a proper truncated log-uniform prior over the noise and truncated log-normal variational approximation that ensures that the KL-term in the evidence lower bound is computed in closed-form. The model leads to structured sparsity by removing elements with a low SNR from the computation graph and provides significant acceleration on a number of deep neural architectures. The model is very easy to implement as it only corresponds to the addition of one dropout-like layer in computation graph.	
06.10.17	Дмитрий Кропотов, научный сотрудник, ВМК МГУ Tensor Train Decomposition for Fast Learning in Large Scale Gaussian Process Models Gaussian Process models is a popular Bayesian approach for solving different machine learning problems, including regression, classification and structured prediction. Training full GP model scales cubically with training set size thus preventing efficient learning in case of large datasets. For this case an inducing inputs approach is usually used that scales linearly with training set size and cubically with number of inducing inputs. Empirical evaluation shows that ability to work with quite a small number of inducing inputs leads to poor performance of GP models in case of large number of features. In this talk, we discuss a learning procedure for GP models that allows using much larger number of inducing inputs. This procedure can be interpreted as a fast variational inference scheme with several approximations made for variational distribution. One of them uses Tensor Train format – a popular approach for compact storing and fast operating with multidimensional tensors.	arxiv, слайды, видео
13.10.17	Naftali Tishbi, Professor, The Hebrew University of Jerusalem Information Theory of Deep Learning ~~Reinforcement Learning under Information Constraints~~ ~~Planning under uncertainty, in the model of the world, noisy observations, or partial control, can be formulated as a tradeoff between control and information flow optimization. In this talk I will present the formulation of this tradeoff, as a generalized Bellman like optimal control problem, an as a general framework for quantifying perception-action cycles in cognitive science. I will present this framework both as a way to study individual control problems under uncertainty and ensemble behavior of a large population. Such stochastic control problems become very important for the analysis of large scale navigation, autonomous vehicles, or flocks of birds and insects. Some recent algorithms derived using this approach will be discussed.~~ https://arxiv.org/abs/1503.02406 https://arxiv.org/abs/1703.00810 Спецсеминар пройдёт в зале Экстрополис	слайды, видео - с предыдущего доклада в Яндексе
20.10.17	Семинара не будет	
27.10.17	Артём Соболев, AI Engineer at Luka Inc. Stochastic Computation Graphs (parts 0, 1 and 2) Deep Neural Networks are known to be very powerful function approximators. Combining DNNs with probabilistic modeling with latent variables has proven to be bilaterally beneficial: one could enrich deep models with stochastic control, train generative models, perform approximate inference, navigate RL agents in uncertain environments. The goal of this talk is to familiarize the audience with the latest advancements in the area. It is assumed the audience has some familiarity with the problem, hence basic concepts will be introduced only briefly to set up the notation. Then we'll move on to the core problem of stochasticity in computational graphs: backpropagation through randomness. We'll discuss general approach and its weaknesses, special cases for continuous and discrete random variables.	слайды, видео
03.11.17	Артём Соболев, AI Engineer at Luka Inc. Stochastic Computation Graphs (part 3).	слайды, видео

10.11.17	Артем Артемов, руководитель компании «Когнитивные системы» Informational Neurobayesian Approach to Neural Networks Training. Opportunities and Prospects A study of the classification problem in context of information theory is presented in the paper. Current research in that field is focused on optimisation and bayesian approach. Although that gives satisfying results, they require a vast amount of data and computations to train on. Authors propose a new concept named Informational Neurobayesian Approach (INA), which allows to solve the same problems, but requires significantly less training data as well as computational power. Experiments were conducted to compare its performance with the traditional one and the results showed that capacity of the INA is quite promising. https://arxiv.org/abs/1710.07264	слайды
17.11.17	Алексей Умнов, младший научный сотрудник ФКН ВШЭ Bayesian Methods in Generative Adversarial Networks Generative Adversarial Networks (GAN) is an effective and actively developing approach for building generative models. GANs still have many unresolved problems, such as being unstable and hard to train, and the mode-collapse problem (missing modes of the distribution and producing non-diverse samples). In my talk I will present you two GAN architectures (BayesianGAN and AlphaGAN) that suggest ways of solving these problems using Bayesian methods. BayesianGAN paper presents the architecture for probability inference for generator parameters, and then uses a distribution over networks rather than one network (which improves the samples diversity). AlphaGAN paper merges the Variational AutoEncoder (VAE) with GAN in order to use the benefits of both architectures	слайды, видео
24.11.17	Максим Кретов, научный сотрудник лаборатории нейронных сетей и глубокого обучения МФТИ Using stochastic computational graphs formalism for optimization of sequence-to-sequence model Variety of machine learning problems can be formulated as an optimization task for some (surrogate) loss function. Calculation of loss function can be viewed in terms of stochastic computational graphs (SCG). We use this formalism to analyze a problem of optimization of famous sequence-to-sequence model with attention and propose reformulation of the task. Examples are given for machine translation (MT). Our work provides a unified view on different optimization approaches for sequence-to-sequence models and could help researchers in developing new network architectures with embedded stochastic nodes. https://arxiv.org/abs/1711.07724 Differentiable lower bound for expected BLEU score In natural language processing tasks performance of the models is often measured with some non-differentiable metric, such as BLEU score. To use efficient gradient-based methods for optimization, it is a common workaround to optimize some surrogate loss function. This approach is effective if optimization of such loss also results in improving target metric. The corresponding problem is referred to as loss-evaluation mismatch. In the present work we propose a method for calculation of differentiable lower bound of expected BLEU score that does not involve computationally expensive sampling procedure such as the one required when using REINFORCE rule from reinforcement learning (RL) framework. Derived lower bound is tight in the sense that for degenerate distributions of candidate text it coincides with exact BLEU score, thus it is fair to refer to this lower bound as "differentiable BLEU score". https://arxiv.org/abs/1712.04708	слайды, видео
01.12.17	Vladimir Kolmogorov, Professor, IST Austria Valued Constraint Satisfaction Problems I will consider the Valued Constraint Satisfaction Problem (VCSP), whose goal is to minimize a sum of local terms where each term comes from a fixed set of functions (called a "language") over a fixed discrete domain. I will present recent results characterizing languages that can be solved using the basic LP relaxation. This includes languages consisting of submodular functions, as well as their generalizations. One of such generalizations is k-submodular functions. In the second part of the talk I will present an application	видео (похожий доклад в ВШЭ) видео и слайды

	of such functions in computer vision. Based on joint work with Igor Gridchyn, Andrei Krokhin, Michal Rolínek, Johann Thapper and Stanislav Živný: http://pub.ist.ac.at/~vnk/papers/BLP-JOURNAL.html http://pub.ist.ac.at/~vnk/papers/VCSP.html http://pub.ist.ac.at/~vnk/papers/POTTS.htm Спецсеминар пройдёт в зале Экстрополис совместно с семинаром Яндекса: https://events.yandex.ru/events/science-seminars/1-dec-2017/	
08.12.17	Спецсеминара не будет (NIPS)	

Весенний семестр 2017 г.

В весеннем семестре 2017 г. спецсеминар проходит в ШАД по пятницам, начало в 18-45.
Для получения пропуска заранее напишите фамилию, имя и отчество Михаилу Фигурнову (michael@figurnov.ru). Язык спецсеминара — английский.

Рассылка спецсеминара, Видео, Страница спецсеминара на machinelearning.ru (архив докладов)

Дата	Докладчик и тема	Материалы
03.02.17	Дмитрий Ветров, Ольга Скороходова, Святослав Яковишин Арабо-израильские войны	видео
10.02.17	Александр Новиков, научный сотрудник ФКН ВШЭ Normalization propagation There are lots of different normalization techniques in deep learning community: Batch Normalization, Instance Normalization, Weight Normalization, and Normalization Propagation, and one cannot train a network with say 50 layers without one form of normalization or another. In this talk, I will discuss why we need to normalize something in neural networks, cover differences between normalization techniques. I will specifically focus on Normalization Propagation, which claims to work as well as Batch Normalization, but do not estimate any statistics from the data and thus can be beneficial from theoretical point of view and can handle batch size 1 scenario. https://arxiv.org/abs/1603.01431	слайды, видео
17.02.17	Евгений Бурнаев, профессор Сколтеха Гауссовские модели для консолидации данных из разных источников Задача консолидации данных из разных источников является одной из основных в приложениях индустриальной инженерии: например, при построении моделей и методов поиска в информационных сетях можно использовать как надежную информацию о	слайды 1 слайды 2, видео

	предпочтениях пользователя, так и менее надежную (косвенную) информацию о том, насколько похожи те или иные интернет-сайты на интернет-сайты, выделенные пользователем; при построении суррогатных моделей зачастую доступны не только высокоточные данные натурного физического эксперимента, но и менее точные данные компьютерных симуляций, моделирующих тот же самый физический феномен. В докладе предполагается рассказать о подходе к консолидации данных из разных источников, использующем регрессию на основе разноточных гауссовских процессов. Будут представлены строгие результаты о минимаксной ошибке интерполяции разноточных данных и показано, как эти результаты можно использовать в прикладном алгоритме для построения оптимального дизайна экспериментов.	
3.03.17	Михаил Хальман, магистр ФКН ВШЭ Neural Conversational Models People are interested in building a human-like chatbot since the beginning of the computer era. The recent advances in deep learning and generative modelling allow us to make chatbots that learn directly from data, thus eliminating the need for programming thousands of hand-crafted rules and templated responses. This approach is not only cheaper, but also much more scalable. However, the problem is far from being solved. In this talk I will give an overview of the most commonly used models for building neural conversational agents such as seq2seq, HRED and their extensions. Their opportunities and limitations will be discussed.	слайды, видео
10.03.17	Михаил Фигурнов, научный сотрудник ФКН ВШЭ Spatially Adaptive Computation Time for Residual Networks We present a deep learning architecture based on Residual Networks that dynamically adjusts the number of executed layers for regions of an image. This architecture is end-to-end trainable, deterministic, and problem-agnostic. It uses two key components: (1) adaptive computation time mechanism; (2) perforated convolutional layer. We present experimental results on ImageNet classification and COCO object detection datasets demonstrating that this architecture improves the computational efficiency of Residual Networks, especially for the higher-resolution images. Then, we demonstrate that the computation time per region correlates well with the human eye fixation positions. Finally, we discuss several ways to extend the presented work. https://arxiv.org/abs/1612.02297	слайды, видео
17.03.17	Александр Чистяков, Лаборатория Касперского Обучение представлений для поведенческих логов и детектирование вредоносной активности на их основе Доклад будет состоять из двух частей: В первой половине доклада будет представлен новый подход для построения признакового описания логируемых данных. Предлагаемый подход основан на построении поведенческого графа специального вида, устойчивого к вариациям в поведении логируемого объекта, и дальнейшем эффективном преобразовании полученного графа в вектор вещественных чисел, описывающий наблюдаемые в логе шаблоны поведения (статья). Вторая половина доклада будет посвящена особенностям задачи обнаружения вредоносных действий в логе системных событий. Мы увидим какие проблемы возникают на практике при классификации изменяющегося во времени объекта (такого как системный лог); рассмотрим подходы, позволяющие обучать классификатор, решающий данные проблемы; и получим удобный механизм для автоматического обеспечения интерпретируемости вердикта модели и для эффективного исправления ложных срабатываний обученного детектора.	слайды, видео
24.03.17	Павел Филонов, Андрей Лаврентьев, Артем Воронцов; Лаборатория Касперского Multivariate Industrial Time Series with Cyber-Attack Simulation: Fault Detection Using an LSTM (доклад на русском языке)	слайды, видео

	One area that strongly requires a technique for multivariate time series analysis is cyber-security for industrial processes. Deep packet inspection (DPI) tool monitors network protocols and provides visibility of sensor and command values inside technological signals represented as a multivariate time series. We adopted an approach based on an LSTM neural network to monitor and detect faults in industrial multivariate time series data. To validate the approach we created a Modelica model of part of a real gasoil plant. By introducing hacks into the logic of the Modelica model, we were able to generate both the roots and causes of fault behaviour in the plant. Having a self-consistent data set with labelled faults, we used an LSTM architecture with a forecasting error threshold to obtain precision and recall quality metrics. The dependency of the quality metric on the threshold level is considered. An appropriate mechanism such as "one handle" was introduced for filtering faults that are outside of the plant operator field of interest. https://arxiv.org/abs/1612.06676	
31.03.17	Арсений Ашуха, магистр МФТИ Variational Dropout Sparsifies Deep Neural Networks Variational dropout is a recent method to learn optimal dropout rates for a neural network in a Bayesian way. During my talk, I will tell about my work with Dmitry Molchanov and Dmitry Vetrov. We extend Variational Dropout to the case when dropout rates are unbounded. Interestingly, it leads to extremely sparse solutions both in fully-connected and convolutional layers. This effect is similar to automatic relevance determination effect in empirical Bayes but has some advantages. Finally, we will discuss future research ways in the area of variational dropout. https://arxiv.org/abs/1701.05369	слайды, видео
07.04.17	Тимур Гарипов, студент ВМК МГУ Successor Representation for Reinforcement Learning Many reinforcement learning algorithms are based on estimation of value functions. There are two widely used approaches to learning value functions: ● model-based algorithms; ● model-free algorithms. However, there is an alternative approach based on the successor representations (SR). The main concept of SR-based algorithms is the estimation of value functions by learning expected representation of states that will be encountered in future. In this talk I will give an overview of the SR idea and the very interesting way to combine it with deep learning. https://arxiv.org/pdf/1606.02396.pdf	слайды, видео 1, видео 2
14.04.17	Maurizio Fillippone, Assistant Professor, EURECOM Practical and Scalable Inference for Deep Gaussian Processes The study of complex phenomena through the analysis of data often requires us to make assumptions about the underlying dynamics. In modern applications, for many systems of interest we are facing the challenge of doing so when very little is known about their mechanistic description. Even when a mechanistic description is available, simulating such systems is so computationally expensive that we cannot use it effectively. Probabilistic models based on Deep Gaussian Processes (DGPs) offer attractive tools to tackle these challenges in a principled way and to allow for a sound quantification of uncertainty. However, inference for DGPs poses huge computational challenges that arguably hinder their wide adoption. In this talk, I will present our contribution to the development of practical and scalable inference for DGPs, which can exploit distributed and GPU computing. In particular, I will introduce a novel formulation of DGPs based on random features that we infer using stochastic variational inference. Through a series of experiments, I will illustrate how our proposal enables scalable deep probabilistic nonparametric modeling and significantly advances the state-of-the-art on inference methods for DGPs.	слайды, видео

21.04.17	Виктор Януш, 3 курс "Learnable optimization strategies using recurrent neural networks" Никита Романов, 5 курс "Normalization Propagation for Deep Neural Networks" Надежда Чиркова, 5 курс "Variational Dropout for Recurrent Neural Networks" Олег Иванов, 5 курс "Missing Features Imputation using Conditional Variational Autoencoders"	видео
28.04.17	Тимур Гарипов, 4 курс "Тензоризованные нейронные сети" Павел Измаилов, 4 курс "Алгоритмы обучения гауссовских процессов для больших объемов данных" Даниил Полыковский, 4 курс "Механизмы внимания в нейронных сетях" Павел Темирчев, 6 курс "Использование нейросетевого подхода для аппроксимации одной гидродинамической модели"	видео
05.05.17	*Не будет*	
12.05.17	Андрей Атанов, 3 курс ФКН ВШЭ "Ensemble distillation" Артем Гадецкий, 3 курс ФКН ВШЭ "Conditional Generators of Words Definitions" Полина Кириченко, 3 курс ФКН ВШЭ "Dealing with the vanishing and exploding gradient problems in recurrent neural networks"	видео
19.05.17	*Не будет*	
26.05.17	Novi Quadrianto, Academic Supervisor, International Laboratory of Deep Learning and Bayesian Methods, Faculty of Computer Science, Higher School of Economics; Assistant Professor, University of Sussex The Privileged Cube in Machine Learning In standard discriminative machine learning models, the assumption is that all features that are being used at training time are available for future data (at deployment time). This assumption however does not always hold. Some features are not available. Some features are too costly in terms of time and money. 3D data are not easily available at deployment time; discarded features from filter/wrapper feature selection methods are not available at deployment time; confidence in crowdsourced annotations is only available at training time. In this talk, I shall review recent approaches to utilize this so-called privileged information in discriminative models. While those applications are important, the full potential of privileged learning has not yet been explored, both in theory and in practice. Therefore, I shall touch upon the Privileged Cube that means the privileged learning paradigm in three dimensions: models (non-Bayesian v. Bayesian), learning problems (non-structured v. structured), and constraints (cost-effective v. transparency). Learning in the Privileged Cube will allow for example making the deployed system to operate with interpretable features while using complex un-interpretable deep features as privileged data at training time.	видео
02.06.17	Pre-defence of Skoltech Master students Valentin Sytov "Deep Neural Descriptors for Image Retrieval" Ekaterina Yakovleva "Multimodal distributions in variational autoencoders" Ilia Yakubovsky "Variational Multilingual Model for Sentence Embedding"	

Осенний семестр 2016 г.

В осеннем семестре 2016 г. спецсеминар проходит в ШАД по пятницам, начало в 18-45.
Для получения пропуска заранее напишите фамилию, имя и отчество Михаилу Фигурнову (michael@figurnov.ru). Язык спецсеминара — английский.

Рассылка спецсеминара, Видео, Страница спецсеминара на machinelearning.ru (архив докладов)

Дата	Докладчик и тема	Материалы
2.09.16	Дмитрий Молчанов, магистр Сколтеха Variational Dropout for Deep Neural Networks and Linear Model Variational dropout is a recent method to learn optimal dropout rates for a neural network in a bayesian way. During my talk I will tell about my work with Arseniy Ashuha. Our research was focused on automated learning of dropout rates. We studied this approach on deep neural networks. Also we applied this approach to linear models and used it for feature selection in a way, similar to automatic relevance determination in RVM.	видео, слайды
9.09.16	Спецсеминар отменяется из-за выступления Кристофа Ламперта 8 сентября.	
16.09.16	Сергей Бартунов, внешний совместитель ФКН ВШЭ One-shot generative modelling There are many existing approaches to generative modelling which appeared recently such as variational autoencoders or adversarial networks. However, most state of the art models are able to produce good results (in terms of visual quality or likelihood) only after extensive training on large datasets. This talk will cover an emerging trend in generative modelling which is often referred to as one-shot learning, i.e. the ability to learn only on several training examples. In addition, a draft of the new model that can generalize over different classes of training examples will be presented.	видео, слайды
23.09.16	Кирилл Струминский, аспирант ФКН ВШЭ Discrete variational autoencoders Datasets composed of discrete classes can be naturally captured by probabilistic models with discrete latent variables. However, usually they are hard to train on large datasets due to a number of reasons. A novel class of probabilistic models, comprising discrete and continuous latent variables will be introduced. Specifically, in these models discrete component captures the distribution over the disconnected smooth manifolds induced by the continuous component. What is more, models in this class can be trained by an extension of variational autoencoder framework. https://arxiv.org/abs/1609.02200	видео, слайды
30.09.16	Артём Гадецкий, студент ФКН ВШЭ	видео, слайды

	AlphaGo or how Deepmind taught machine to win The game of Go has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves. During my talk I will tell about how Deepmind used convolutional neural networks for reducing effective search space and how networks were combined with Monte Carlo Tree Search. http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html https://arxiv.org/abs/1412.6564	
7.10.16	Дмитрий Кропотов, научный сотрудник ВМК МГУ Optimizing Neural Nets using Kronecker-factored Approximate Curvature (K-FAC) In classic optimization, second-order methods and their variants (Hessian-free Newton, natural gradient, L-BFGS and others) forms a state-of-the-art approach - they do not need any user-tunable parameters and can outperform by far simple strategies like gradient descent. However, application of these methods in stochastic and non-convex setting (most notable example is learning neural nets) remains a very challenging problem. Numerous attempts in this field (e.g. stochastic L-BFGS, Hessian-free for deep learning) hasn't led to successful and popular algorithm, and thus many practitioners still prefer using here plain stochastic gradient descent (SGD) and its simple modifications. Recently, a new second-order optimization method with Kronecker-factored approximate Fisher matrix (K-FAC) has been proposed. One of the key advantages of this method is its high-quality approximation for full Fisher matrix based on information from stochastic mini-batches. The iteration complexity and memory cost of the method is only a constant factor higher comparing to SGD. However, thanks to using second-order information, in practice the new method requires several orders less iterations for convergence and has no problem-specific parameters. In my talk, I tell about basics of natural gradient approach, K-FAC approximation ideas and introduce the resulting algorithm for optimization of fully connected neural networks. Besides, I give a glimpse on KFC approximation - a recent modification of these method for convolutional networks.	видео, слайды
14.10.16	Дмитрий Ульянов, аспирант Сколтеха Image artistic style transfer, neural doodles and texture synthesis A recent advances in image style transfer allowed incredible end-user applications. At first, Gatys et al. demonstrated that deep neural networks can generate beautiful textures and stylized images from a single example. The core idea of the method was used then to create so-called neural doodles. While the visual quality of both style transfer and neural doodles was astonishing, the methods required a slow and memory-consuming optimization process, which limited their usage. We lately improved the speed of both algorithms significantly, while preserving the quality. This allowed almost real-time stylization using GPU, the method was used as a core technology in several successful applications. In this talk we overview and discuss the mentioned algorithms.	видео, слайды
21.10.16	Влад Шахуро, аспирант ФКН ВШЭ Training generative neural networks using Maximum Mean Discrepancy There are several approaches to training generative models based on neural networks. The most popular are variational autoencoder and adversarial networks. In this talk I tell about alternative approach for training generative models. It is based on technique from statistical hypothesis testing known as maximum mean discrepancy (MMD). Such technique leads to a simple loss function that tries to match all orders of statistics between training dataset and samples from the model which can be trained by backpropagation. Compared to GAN, training with MMD loss function is easier. One doesn't have to design a discriminator and no tricky alternating training procedure is required. https://arxiv.org/pdf/1502.02761v1.pdf https://arxiv.org/pdf/1505.03906v1.pdf https://arxiv.org/pdf/1606.02556v4.pdf	видео, слайды
28.10.16	Владимир Спокойный, WIAS and Humboldt University Berlin	видео, слайды

	Clustering using adaptive weights The paper discusses a new method of unsupervised learning for high dimensional data based on the idea of adaptive weights from Polzehl and Spokoiny (2000). The procedure recovers the unknown clustering structure without any prior information about the number of clusters, their size, distance between clusters, etc.The approach extends the popular k-mean and density based clustering procedures by using dynamically updated local weights. Theoretical results describe two major features of the method: propagation within a homogeneous region and separation between two different regions. Numerical results show state-of-art performance of the new procedure.	
4.11.16	Выходной	
11.11.16	Александр Гасников, МФТИ Современные численные методы стохастической оптимизации и их приложения В докладе пойдет речь о том как с помощью стохастической оптимизации можно решать задачи математической статистики (агрегирование оценок) и статистической теории обучения. Доклад будет носить обзорный характер.	видео, слайдгы
18.11.16	В.С. Смолин, ИПМ им. М.В. Келдыша РАН Табличное представление и интерполирование функций нейросетями Предлагается несколько идей математического описания как самих сетей формальных нейронов, так и выполняемых в них преобразований, позволяющих лучше понять характер обработки информации в нейросетях. Проводится аналогия между табличной функцией и нейросетью. Работы по развитию универсальной аппроксимационной теоремы для функций многих переменных ведутся давно, начиная с работ Колмогорова и Арнольда в 1956-7 годах, получен ряд интересных результатов, но широкого применения нейросетевые аппроксиматоры пока не получили. Это связано с недостаточной разработкой таких вопросов аппроксимации функций многих переменных, как проклятие размерности, преобразование функций к монотонному виду, задание целевых функций и одноэкстремальность адаптации, масштаб прогнозирования. Обсуждаются возможные пути решения вышеперечисленных и некоторых связанных с ними вопросов и на основе рассмотрения делаются предложения об использовании аппроксимирующих нейросетей для решения широкого круга прикладных задач.	видео, слайды
25.11.16	Александр Панин, Yandex Data Factory Variational Information Maximizing Exploration When it comes to solving practical problems, performance of reinforcement learning algorithms usually depends highly on efficient environment exploration. However, classical exploration strategies (e-greedy, boltzmann) have several common drawbacks that jeopardize training speed. Informally, if you want to learn to program in java, having already learned python, randomly mistyping 10% of characters (e-greedy) and keeping those that compiled will likely yield poor results. We'd like to describe a method devised by Abeel et. al that tackles this problem by means of variational inference.	видео, слайды
2.12.16	Спецсеминар отменяется из-за выступления Питера Рихтарика в Яндексе.	
9.12.16	Все ушли на NIPS...	
16.12.16	Даниил Полыковский, студент ВМК МГУ Neural networks. *from:* LSTM, *to:* Neural Computer	видео, слайды

	Some essential problems (like sequence sorting, copying and reversal) can not be solved with vanilla LSTM networks, however standard algorithms and data structures can do that. Fusion of these two worlds can expand a class of solvable problems. ML backed up with data structures is one first steps forward to that vision. Recently proposed Neural Turing Machines, Neural Data Structures and Differentiable Neural Computers are currently one of candidates for a soon break-through. For example, problems with structured data such as finding the shortest path can be solved with these techniques.	
23.12.16	Зачёт по спецсеминару (для студентов группы).	
27.12.16	Новогодний коллоквиум по статьям с NIPS 2016 Коллоквиум пройдёт на ФКН ВШЭ в ауд. 205 с 15:10 до 19:00. Список статей см. ниже	
30.12.16	Пересдача зачёта по спецсеминару (для студентов группы).	

Статьи на новогоднем коллоквиуме: видео часть 1 часть 2

Докладчик	Статья	Ссылка
Ашуха	Learning Structured Sparsity in Deep Neural Networks	https://papers.nips.cc/paper/6504-learning-structured-sparsity-in-deep-neural-networks
Ветров	The Generalized Reparameterization Gradient	http://papers.nips.cc/paper/6328-the-generalized-reparameterization-gradient.pdf
~~Гарипов~~ Новиков	Residual Networks Behave Like Ensembles of Relatively Shallow Networks	https://arxiv.org/abs/1605.06431
Игнатов Д.И.	Something interesting from tensor-learn.org workshop @ NIPS	
Лобачева	Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks	https://arxiv.org/abs/1602.07868
Молчанов	Variance Reduction in Stochastic Gradient Langevin Dynamics	http://papers.nips.cc/paper/6293-variance-reduction-in-stochastic-gradient-langevin-dynamics.pdf
Неклюдов	Boosting Variational Inference	https://arxiv.org/pdf/1611.05559v1.pdf
Новиков	Using fast weights to attend to the recent past	https://arxiv.org/abs/1610.06258
Подоприхин	Learning to Play in a Day: Faster Deep Reinforcement Learning by Optimality Tightening	https://arxiv.org/abs/1611.01606
Соколов	Sequential Neural Models with Stochastic Layers	https://arxiv.org/pdf/1605.07571v2.pdf
Струминский	Operator Variational Inference/Stein Variational Gradient Descent	https://arxiv.org/pdf/1610.09033v2.pdf
Фигурнов	Doubly Convolutional Neural Networks	https://arxiv.org/abs/1610.09716