Инфа ниже всё ещё актуальна, но гайд больше не поддерживается и поддерживаться не будет

Текст, скорее всего, будет еще дополняться, по всем вопросам и предложениям писать в Telegram: @STNLk За книги по Питону и курс CSC спасибо Валере.

Начну с того, что на освоение всего с нуля (под нулем я понимаю школьное знание математики) уйдет достаточно много времени. По самым оптимистичным оценкам 2+ лет. Тут, конечно, имеется в виду освоение на уровне приемлемом, чтобы пойти куда-то джуном и понимать содержание последних статей по сабжу. Если такое не устраивает, есть целый интернет с сотнями курсов "Как освоить мошинлернинг за микросек без смс". Также надо добавить, что очень желательно уметь читать литературу и воспринимать лекции на английском. Большое число годноты на русский не переведено и переведено в ближайшее время не будет.

Не претендую на полноту гайда и вообще его адекватность. Я сам шарю тему, наверное, на уровне где-то между джуном и миддлом, ближе к миддлу. Это скорее более похоже на повторение моего пути.

Разобью все на три раздела: математичная база, непосредственно машинлернинг + сопутствующие прикладные скиллы с небольшим уклоном в диплернинг в computer vision (далее CV) в силу того, что я сам на нем специализируюсь.

TL;DR

Для тех, кто математику уже знает либо учить не хочет, порядок такой: Пропускаем раздел с математикой, сразу идем в машинлернинг. Если умеем в питон, то пропускаем пункт 0, изучаем 1-а, 1-b, 2-а, 2-b, 2-с, получаем 300к/наносек. Далее уже сами пробежитесь по списку и подумаете, что надо подтянуть.

МАТЕМАТИКА

Ну тут все ясно, без матеши никуда. Каждый пункт из этого раздела заменяется соответствующим университетским курсом.

- 1. Всякая разная дискретка, ликбез для тех, кто вообще никогда не сталкивался с вузовской математикой не умеет в комбинаторику, впервые слышит про графы и так далее и тому подобное.
 - а. Два первых раздела из трех этой книги http://rubtsov.su/public/hse/2017/DM-HSE-Draft.pdf покрывают все необходимое. Можно просто пройтись по оглавлению и прочитать параграфы с незнакомыми названиями. Читается легко, пререквизитов не требует.
 - b. Лекции Райгородского по комбинаторике https://mipt.lectoriy.ru/course/Maths-CombinatoricsBasics-L15/lectures, семинары к ним https://mipt.lectoriy.ru/course/Maths-CombinatoricsBasics-15S. Входим в катарсис. То же самое, только на курсере https://www.coursera.org/learn/kombinatorika-dlya-nachinayushchikh, чуть более объемный курс с курсеры https://www.coursera.org/learn/modern-combinatorics. Райгородский читает лекции так, что понятно будет и дошкольнику, однако ссылки, которые я привел выше, ограничиваются только комбинаторикой. У него есть большой курс по теории графов на курсере и более полные курсы по комбинаторике и теории чисел, но я думаю, что сюда их пихать все незачем
- 2. Математический анализ. Основа основ. Конкретно для наших целей достаточно будет первого года стандартного вузовского курса матана. А именно, всё до дифференциального исчисления нескольких переменных включительно. Матан я настоятельно рекомендую основательно изучить в вузе. Самостоятельно погрузиться в него по книжкам и курсам довольно тяжело, но я все же приведу некоторый список.
 - а. Зорич В. А. Математический анализ. Часть 1. Прочтение и осмысление его эквивалентно плотному изучению матана на математической специальности хорошего вуза и даже больше. Пожму крепко руку человеку, который с нуля изучил матан только по этому учебнику.

- b. Хорошие и понятные физтеховские лекции от Иванова Г.Е. https://mipt.lectoriy.ru/course/Maths-Basic_Analysis-2sem
- с. Лекции чебышатника от Кислякова С. https://www.youtube.com/playlist?list=PL-_cKNuVAYAU1iqXkA388zr0c45 https://www.youtube.com/playlist?list=PL-_cKNuVAYAU1iqXkA388zr0c45 https://www.youtube.com/playlist?list=PL-_cKNuVAYAU1iqXkA388zr0c45 https://www.youtube.com/playlist?list=PL-_cKNuVAYAU1iqXkA388zr0c45 https://www.youtube.com/playlist?list=PL-_cKNuVAYAU1iqXkA388zr0c45
- 3. Линейная алгебра. Один из самых важных предметов наряду с матаном, однако линал гораздо проще в изучении, и изучить его вполне возможно по книгам и курсам самостоятельно. К сожалению, хороших книг по линалу я не знаю. Могу посоветовать неплохие курсы
 - а. Курс на английском. Лютейшая годнота от топового препода H.Гусева https://stepik.org/course/79/syllabus.
 - b. Физтеховские лекции от Кожевникова П.А. Читает неплохо. Однако курс линала на физтехе является логическим продолжением курса аналитической геометрии, поэтому порой будут требоваться какие-то знания из неё (алгебра матриц и детерминанты входит в программу ангема, например) https://mipt.lectoriy.ru/course/LinearAlgebra. Ангем, оттуда достаточно последних 3-4 лекций https://mipt.lectoriy.ru/course/Math-analyth_geom/lectures.
- 4. Теория вероятностей. Без матана и линала можно изучить разве что какие-нибудь детские версии теорвера, вроде введения в теорвер от Райгородского https://mipt.lectoriy.ru/course/Maths-ProbabilityTheoryBasics-L15. Для совсем плотного изучения теорвера неплохо бы основы теории меры понять. Часто теорию меры вводят в курсах матана. Довольно годные заметки к лекциям Н.Гусева https://ngusev.ru/mt/mt-lectures.pdf (это уже для любителей обмазаться матешей).
 - а. Хорошая книга. Боровков А. А. Теория вероятностей. Оттуда достаточно первых 8 глав (до главы с предельными теоремами включительно). Будет полезна также глава 13 про теорию информации и глава 20 про марковские цепи. В приложении есть введение во все нужное, связанное с теорией меры.
 - b. Неплохой сжатый курс https://stepik.org/course/3089/syllabus
- 5. Математическая статистика. Изучаем после теорвера. Тут я приведу только один источник, потому что сам кроме того, что было в вузе, плотно изучал только эту книжку. Лагутин М. Б. Наглядная математическая статистика.

- Книжка определенно хороша. Есть куча задач, и все они снабжены авторскими решениями.
- 6. Оптимизация. Без матана и линала нечего лезть. Чрезвычайно важная штука для машинки, однако вполне можно обходиться только базовыми понятиями из нее, а остальное уже изучать по ходу дела. На всякий случай приведу тут две книжки, после прочтения которых можно стать богом оптимизации. Читать лучше в том порядке, в котором они приведены.
 - a. Boyd S., Boyd S. P., Vandenberghe L. Convex optimization. Классное введение в выпуклую оптимизацию (на 700 страниц, ага).
 - b. Nesterov Y. Lectures on convex optimization. Более продвинутая книга, написанная нашим соотечественником. Не сильно ошибусь, если скажу, что в оптимизации он сейчас самый видный челибос.
- 7. Прочее: теория информации, графы, алгоритмы, чего-то еще, что я забыл. Все это не стоит подробного описания и изучения, как мне кажется. Все важные математические вещи, которые не входят в основное содержание предметов выше, описаны либо там в отдельных параграфах или приложениях, либо в книгах и курсах уже по непосредственно машинному обучению. На крайняк сами нагуглите требуемую инфу.



Машин Лернинг

- 0. Python. Без питона в машинке вы сами знаете кто. Базовый питон с нулевого уровня:
 - а. https://pythontutor.ru/ куча задачек, пацанский курс
 - b. https://stepik.org/course/512/promo на степике куча курсов по питону

Питон продвинутый

a. https://www.youtube.com/playlist?list=PLlb7e2G7aSpTTNp7HBYzCByaE1h54ruW годнота от CSC. Если более свежие версии этого курса на ютубе, однако тут лектор поинтереснее рассказывает.

b. Fluent Python: Clear, Concise, and Effective Programming. Эту книгу я ещё не начинал читать, но её советовал Валера.

Питон для машинлернеров. Тут будут скорее туториалы по полезным библиотекам для машинлернистов. Описывать сами библиотеки я не буду. Туториалы можно читать уже параллельно с курсами непосредственно по машинке (как и материалы по продвинутому питону) либо обращаться к ним только по мере надобности.

- a. http://scipy-lectures.org/index.html хороший туториал по питру, scipy, и ещё по некоторым полезным либам понемножку.
- b. https://pandas.pydata.org/pandas-docs/stable/getting_started/tutorials.ht ml официальные туториалы по pandas
- с. Pytorch популярная библиотека для глубокого обучения. https://pytorch.org/tutorials/ официальные туториалы. Почему pytorch, а не tensorflow или какие-нибудь его производные, например? Просто потому что я сам только в него и умею, и этого мне пока что достаточно.

1. Наконец, машинное обучение.

- a. Начнем с лютейшей годноты. https://www.coursera.org/specializations/machine-learning-data-analysi
 - §. Специализация от Яндекса и МФТИ. После ее прохождения можно смело идти на собесы на джуна. Параллельно с первым курсом специализации надо проходить этот курс со степика https://stepik.org/course/3356/syllabus. Он задумывался как дополнение к первому курсу и рекомендован авторами специализации к прохождению. Там больше тематических задачек и материалов по питону со всеми вытекающими. После 2-3 курса специализации я уже рекомендую поглядеть и порешать учебные соревнования на Kaggle. Практика в этом деле крепко забустит понимание. После окончания специализации уже можно пробовать силы в действующих соревнованиях на Kaggle, собирать портфолио, так сказатб.
- b. Отличный курс Нейчева, который читается сейчас на фивте и фупме. Он состоит из двух частей. Вторая часть больше к диплернингу относится, поэтому я отнесу ее в другой раздел. А

первую часть можно смотреть либо параллельно со специализацией, либо сразу после нее ради закрепления https://www.youtube.com/playlist?list=PL4_hYwCyhAvZyW6qS58x4uElZgAkMVUvj. К курсу есть семинары, на которых пишется и разбирается код

https://www.youtube.com/playlist?list=PL4_hYwCyhAvYPOWn6e44
RKxEfRWEsPA1z. Репозиторий курса
https://github.com/ml-mipt/ml-mipt/tree/master

- c. Friedman J., Hastie T., Tibshirani R. The elements of statistical learning. Классика. Машинное обучение со статистической точки зрения.
- d. Bishop C. M. Pattern recognition and machine learning. Я ее еще не читал, но мне ее рекомендовал мой преподаватель и коллега Писов М. Посложнее, чем книга из предыдущего пункта.
- e. Shalev-Shwartz S., Ben-David S. Understanding machine learning: From theory to algorithms. Отличный выбор для тех, кто хочет обмазаться доказательствами и теоремками.
- 2. Нейросеточки/глубокое обучение/диплернинг/итд. Порядок тут особого значения не имеет, если не оговорено иного в самом пункте. Наиболее быстрый способ вката это прохождение пункта 2-с. Предварительное изучение пунктов 2-а и 2-b может этот вкат облегчить. Пункты же 2-d и ниже нужны, скорее, чтобы углубить какие-то знания или вспомнить благополучно забытое.
 - a. https://stepik.org/course/50352/ Отличный курс от Samsung по сеточкам в CV. И вообще хороший вводный курс по диплернингу. Не очень много практики, подойдет, скорее, для того, чтобы осознать, как все примерно работает.
 - b. https://stepik.org/course/54098/syllabus Курс Samsung по Natural Language Processing (NLP). Рекомендуется проходить после предыдущего пункта.
 - с. Вторая часть курса Нейчева. Первая была в 1-с. Лекции https://www.youtube.com/playlist?list=PL4_hYwCyhAvZeq93ssEUaR47xhvs7IhJM. Семинары https://www.youtube.com/playlist?list=PL4_hYwCyhAvYvuHz_PKIEV-kOsK2bwUBg. Курс является также неплохим обзорным и

- вводным курсом по нейросетям. При этом затрагиваются довольно свежие темы, публикации по которым выходили в свет в последние 3-7 лет. В курсе читаются NLP, Reinforcement Learning и CV, именно в таком порядке. Если решать домашки (их не очень много) и следить за семинарами, то можно и на одном только этом курсе выехать на уровень крепкого джуна.
- d. https://www.deeplearningbook.org/ Прекрасная вводная книга от автора оригинальной статьи про GANы. В первых главах книги дается в обзорном порядке требуемая математика. Полезно, чтобы быстренько вспомнить некоторые вещи. В целом, книга читается довольно легко, немалый объем текста посвящен описанию мотивации и истории, что значительно упрощает чтение и понимание.
- e. Классика жанра, курсы Stanford University, CS231N (CV) https://www.youtube.com/playlist?list=PLoROMvodv4rOhcuXMZkNm7j3fVwBBY42z. Выше плейлисты с лекциями за предыдущие года. Можно нагуглить их домашки и попытаться прорешать их, параллельно просматривая лекции. Оба курса гораздо более подробно освещают CV и NLP, чем 2-а, 2-b и 2-с (хотя последний пункт вполне себе самодостаточен).
- f. https://d21.ai/ Книга с кодом.