

Syllabus for NLP 243 Machine Learning for Natural Language Processing

Description:

Introduction to machine learning models and algorithms for Natural Language Processing including deep learning approaches. NLP 243 is a course targeted at Professional MS students, which will focus more on applications and current use of these methods in industry. Topics include an introduction to standard neural network learning methods such as feed-forward neural networks, recurrent neural networks, convolutional neural networks, and encoder-decoder models with applications to natural language processing problems such as utterance classification and sequence tagging.

Course goals: The aim of this course is to teach students recent methods in machine learning as applied to problems in natural language processing, with a primary focus on deep learning methods. By the end of the course the students will have experience and skills with applying methods and tools to standard NLP problems such as text classification and sequence to sequence problems such as machine translation.

Learning outcomes: This is a core course for the NLP MS degree. The learning outcomes are based on the PLOs for the NLP MS.

Learning outcomes:

- The ability to propose and apply appropriate existing algorithms and tools in the area of Machine Learning to address questions related to real-life Natural Language Processing problems.
- Proficiency in the use of both traditional machine learning and deep learning algorithms commonly used in Natural Language Processing and corresponding publicly available tools.

Prerequisites: None except entry requirements for the program. This is a core course taught in the fall in the first quarter of the program. Enrollment is restricted to NLP graduate students.

Weekly Schedule: The class will consist of 3 hours and 15 minutes lecture, and an obligatory 1 hour and 45 minute section each week run by a TA for hands on help with concepts and programming assignments. TAs will also hold office hours 2 hours a week.

Textbook/Materials:

This course will use two textbooks, building on the material taught in NLP 243 which is a core course that all students take. The first text explains concepts the theory behind them while the second book provides hands-on practice of concepts for NLP.

- “Dive Into Deep Learning”, Ashton Zhang, Zack C. Lipton, Mu Li, Alex Smola. <http://d2l.ai>.
- “Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning” 1st Edition. Delip Rao and Brian McMahan.
<https://www.amazon.com/Natural-Language-Processing-PyTorch-Applications/dp/1491978236>

Student Hours: Students must attend the lecture which consists of 3 hours and 15 minutes lecture, and also must attend an obligatory 1 hour and 45 minute section each week run by a TA for hands on help with concepts and programming assignments. Programming assignments and projects should take approximately 10 hours per week of independent work.

Week by Week Syllabus:

Week 1: Introduction and overview of machine learning and key concepts, including probability, decision theory, and generative models/discriminative models/discriminants.

Week 2:

- Text Classification and Data Sets, representation of natural language, feature extraction and representation, tools for deriving features of language input.
- Text Classification: Bayesian learning and parameter estimation

Week 3:

- Linear classification and the Perceptron algorithm
- Batch learning: Decision Trees and Artificial Neural Networks

Week 4:

- Word Embeddings, The Skip-Gram Model The Continuous Bag Of Words (CBOW) Model,
 - Training methods for Word Embeddings, Datasets,
 - Approximate training methods, datasets, subword embeddings, Global vectors
 - Tasks for Word Embeddings, Finding Synonyms and Analogies

Week 5:

- Context-aware word embeddings (ELMO, BERT, GPT)
 - Pre-training - Fine-tuning for tasks
- Linear Neural Networks
- Multi-Layer Perceptrons

Week 6:

- Convolutional Neural Networks
- Text Sentiment Classification: Using Convolutional Neural Networks (textCNN)

Week 7:

- Recurrent Neural networks
- Sequence classification using Recurrent Neural Networks

Week 8:

- Recurrent Neural networks
- Sequence classification using Recurrent Neural Networks

Week 9:

- Encoder-Decoder Models, Seq2Seq tasks
- Attention Mechanism
- Encoder/Decoder Example applications: machine translation, NLG, paraphrase generation

Week 10:

- Transformers

Assessments: Requirements include a midterm, final, 3 programming assignments and a project.

Student hours: Students must attend the lecture which consists of 3 hours and 15 minutes lecture, and also must attend an obligatory 1 hour and 45 minute section each week run by a TA for hands on help with concepts and programming assignments. Programming assignments and readings should take approximately 10 hours per week of independent work

Final Grading: The grade will be calculated based on 10% for each programming assignment, 20% on final project, 25% midterm, 25% final.

Support

- **Support for students with disabilities** UC Santa Cruz is committed to creating an academic environment that supports its diverse student body. If you are a student with a disability who requires accommodations to achieve equal access in this course, please submit your Accommodation Authorization Letter from the Disability Resource Center (DRC) to me privately during my office hours or by appointment, preferably within the first two weeks of the quarter. At this time, I would also like us to discuss ways we can ensure your full participation in the course. I encourage all students who may benefit from learning more about DRC services to contact DRC by phone at 831-459-2089 or by email at drc@ucsc.edu.

- **Support for students with other difficulties** While we sincerely hope that you will be able to pursue your studies peacefully and worry-free, we are aware that in some cases difficulties happen that are beyond your control. You should always feel free and comfortable to bring up any problem with the instructor, but if this is not sufficient, or if you prefer professional help, here are several campus resources that you may want to consider contacting:
 - [UC Care](#) which is a confidential space to discuss issues of dating violence, sexual assault and stalking.
 - [Slug Support](#) where you can ask for help on many practical issues, including dealing with a financial crisis, problems with your living situation, computers, books, etc.
 - [CAPS](#), which provides counseling and psychological services to students
 -

- **Title IX reporting disclosure** : Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. If you have experienced sexual harassment or sexual violence, you can receive confidential support and advocacy at the Campus Advocacy Resources and Education (CARE) Office by calling (831) 502-2273. In addition, Counseling and Psychological Services (CAPS) can provide confidential, counseling support, (831) 459-2628. You can also report gender discrimination directly to the University's Title IX Office, (831) 459-2462. Reports to law enforcement can be made to UCPD, (831) 459-2231 ext. 1. For emergencies call 911. Faculty and Teaching Assistants are required under the UC Policy on Sexual Violence and Sexual Harassment to inform the Title IX Office should they become aware that you

or any other student has experienced sexual violence or sexual harassment. If you prefer to speak to someone confidentially, please contact UC Care (see above).

Work ethics

Academic Integrity Academic integrity is the cornerstone of a university education. Academic dishonesty diminishes the university as an institution and all members of the university community. It tarnishes the value of a UCSC degree. All members of the UCSC community have an explicit responsibility to foster an environment of trust, honesty, fairness, respect, and responsibility. All members of the university community are expected to present as their original work only that which is truly their own. Plagiarism of any kind is unacceptable. All members of the community are expected to report observed instances of cheating, plagiarism, and other forms of academic dishonesty in order to ensure that the integrity of scholarship is valued and preserved at UCSC. Any student found in violation of the UCSC Academic Integrity policy may face both academic sanctions imposed by the instructor of record and disciplinary sanctions imposed by the graduate division. Violations of the Academic Integrity policy can result in dismissal from the university and a permanent notation on a student's transcript. For the full policy and disciplinary procedures on academic dishonesty, students and instructors should refer to the [Academic Integrity page](#) at the [Division of Undergraduate Education](#) or [Graduate Division](#).