Note that university employees – including professors and TAs – are required to report what they know about incidents of sexual or relationship violence, stalking and sexual harassment to the Title IX Office. Students can learn more at https://vaden.stanford.edu/sexual-assault (are recovering from).

Schedule

Updated lecture **slides** will be posted here shortly before each lecture. Other links contain last year's slides, which are mostly similar.

Lecture **notes** will be uploaded a few days after most lectures. The notes (which cover approximately the first half of the course content) give supplementary detail beyond the lectures.

Date: Tue Jan 10

Week 1

Description: Word Vectors *(by John Hewitt)*
[slides (slides/cs224n-2023-lecture01-wordvecs1.pdf)] [notes (readings/cs224n_winter2023_lecture1_notes_draft.pdf)]

Gensim word vectors example:
[code (materials/Gensim.zip)] [preview (materials/Gensim%20word%20vector%20visualization.html)]

Suggested Readings:

1. Efficient Estimation of Word Representations in Vector Space (http://arxiv.org/pdf/1301.3781.pdf) (original word2vec paper)
2. Distributed Representations of Words and Phrases and their Compositionality (http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf) (negative sampling paper)

Events: Assignment 1 **out**
[code (assignments/a1.zip)]
[preview (assignments/a1_preview/exploring_word_vectors.html)]

Deadlines:

Date: Thu Jan 12

Description: Word Vectors, Word Window Classification, Language Models
[slides (slides/cs224n-2023-lecture02-wordvecs2.pdf)] [notes (readings/cs224n-2019-notes02-wordvecs2.pdf)]

Suggested Readings:

1. GloVe: Global Vectors for Word Representation (http://nlp.stanford.edu/pubs/glove.pdf) (original GloVe paper)
2. Improving Distributional Similarity with Lessons Learned from Word Embeddings (http://www.aclweb.org/anthology/Q15-1016)
3. Evaluation methods for unsupervised word embeddings (http://www.aclweb.org/anthology/D15-1036)

Additional Readings:

1. A Latent Variable Model Approach to PMI-based Word Embeddings (http://aclweb.org/anthology/Q16-1028)
2. Linear Algebraic Structure of Word Senses, with Applications to Polysemy (https://transacl.org/ojs/index.php/tacl/article/viewFile/1346/320)
3. On the Dimensionality of Word Embedding (https://papers.nips.cc/paper/7368-on-the-dimensionality-of-word-embedding.pdf)

Events:

Deadlines:

Date: Fri Jan 13

Description: Python Review Session
[slides (readings/cs224n-python-review-2023.pdf)] [colab (https://colab.research.google.com/drive/1hxWtr98jXqRDs_rZLZcEmX_hUcpDLq6e?usp=sharing)]

2:30pm - 3:20pm
Gates B03

Events:

Deadlines:

Date: Tue Jan 17

Week 2

Description: Backprop and Neural Networks
[slides (slides/cs224n-2023-lecture03-neuralnets.pdf)] [notes (readings/cs224n-2019-notes03-neuralnets.pdf)]

Suggested Readings:

1. matrix calculus notes (readings/gradient-notes.pdf)
2. Review of differential calculus (readings/review-differential-calculus.pdf)
3. CS231n notes on network architectures (http://cs231n.github.io/neural-networks-1/)
4. CS231n notes on backprop (http://cs231n.github.io/optimization-2/)
5. Derivatives, Backpropagation, and Vectorization (http://cs231n.stanford.edu/handouts/derivatives.pdf)
6. Learning Representations by Backpropagating Errors (http://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf) (seminal Rumelhart et al. backpropagation paper)

Additional Readings:

1. Yes you should understand backprop (https://medium.com/@karpathy/yes-you-should-understand-backprop-e2f06eab496b)
2. Natural Language Processing (Almost) from Scratch (http://www.jmlr.org/papers/volume12/collobert11a/collobert11a.pdf)

Events: Assignment 2 **out**
[code (assignments/a2.zip)]
[handout (assignments/a2.pdf)]
[latex template (assignments/a2_latex_template.zip)]

Deadlines: Assignment 1 **due**

Date: Thu Jan 19

Description: Dependency Parsing
[slides (slides/cs224n-2023-lecture04-dep-parsing.pdf)] [notes (readings/cs224n-2019-notes04-dependencyparsing.pdf)]
[slides (annotated) (slides/cs224n-2021-lecture04-dep-parsing-annotated.pdf)]

Suggested Readings:

1. Incrementality in Deterministic Dependency Parsing (https://www.aclweb.org/anthology/W/W04/W04-0308.pdf)
2. A Fast and Accurate Dependency Parser using Neural Networks (https://www.emnlp2014.org/papers/pdf/EMNLP2014082.pdf)
3. Dependency Parsing (http://www.morganclaypool.com/doi/abs/10.2200/S00169ED1V01Y200901HLT002)
4. Globally Normalized Transition-Based Neural Networks (https://arxiv.org/pdf/1603.06042.pdf)
5. Universal Stanford Dependencies: A cross-linguistic typology (http://nlp.stanford.edu/~manning/papers/USD_LREC14_UD_revision.pdf) (http://nlp.stanford.edu/~manning/papers/USD_LREC14_UD_revision.pdf)
6. (http://nlp.stanford.edu/~manning/papers/USD_LREC14_UD_revision.pdf)Universal Dependencies website (http://universaldependencies.org/)
7. Jurafsky & Martin Chapter 14 (https://web.stanford.edu/~jurafsky/slp3/14.pdf)

Events:

Deadlines:

Date: Fri Jan 20

Description: PyTorch Tutorial Session
[colab notebook (https://colab.research.google.com/drive/13HGy3-uIIy1KD_WFhG4nVrxJC-3nUUkP?usp=sharing)]

3:30pm - 4:20pm
Gates B01

Events:

Deadlines:

Date: Tue Jan 24

Week 3

Description: Recurrent Neural Networks and Language Models
[slides (slides/cs224n-2023-lecture05-rnnlm.pdf)] [notes (lectures 5 and 6) (readings/cs224n-2019-notes05-LM_RNN.pdf)]

Suggested Readings:

1. N-gram Language Models (https://web.stanford.edu/~jurafsky/slp3/3.pdf) (textbook chapter)
2. The Unreasonable Effectiveness of Recurrent Neural Networks (http://karpathy.github.io/2015/05/21/rnn-effectiveness/) (blog post overview)
3. Sequence Modeling: Recurrent and Recursive Neural Nets (http://www.deeplearningbook.org/contents/rnn.html) (Sections 10.1 and 10.2)
4. On Chomsky and the Two Cultures of Statistical Learning (http://norvig.com/chomsky.html)
5. Sequence Modeling: Recurrent and Recursive Neural Nets (http://www.deeplearningbook.org/contents/rnn.html) (Sections 10.3, 10.5, 10.7-10.12)
6. Learning long-term dependencies with gradient descent is difficult (http://ai.dinfo.unifi.it/paolo//ps/tnn-94-gradient.pdf) (one of the original vanishing gradient papers)
7. On the difficulty of training Recurrent Neural Networks (https://arxiv.org/pdf/1211.5063.pdf) (proof of vanishing gradient problem)
8. Vanishing Gradients Jupyter Notebook (https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/lectures/vanishing_grad_example.html) (demo for feedforward networks)
9. Understanding LSTM Networks (http://colah.github.io/posts/2015-08-Understanding-LSTMs/) (blog post overview)

Events: Assignment 3 **out**
[code (assignments/a3.zip)]
[handout (assignments/a3_handout.pdf)]
[latex template (assignments/a3_latex.zip)]

Deadlines: Assignment 2 **due**

Date: Thu Jan 26

Description: Seq2Seq, MT, Subword Models
[slides (slides/cs224n-2023-lecture06-fancy-rnn.pdf)] [notes (lectures 5 and 6) (readings/cs224n-2019-notes05-LM_RNN.pdf)]

Suggested Readings:

1. Statistical Machine Translation slides, CS224n 2015 (https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1162/syllabus.shtml) (lectures 2/3/4)
2. Statistical Machine Translation (https://www.cambridge.org/core/books/statistical-machine-translation/94EADF9F680558E13BE759997553CDE5) (book by Philipp Koehn)
3. BLEU (https://www.aclweb.org/anthology/P02-1040.pdf) (original paper)
4. Sequence to Sequence Learning with Neural Networks (https://arxiv.org/pdf/1409.3215.pdf) (original seq2seq NMT paper)
5. Sequence Transduction with Recurrent Neural Networks (https://arxiv.org/pdf/1211.3711.pdf) (early seq2seq speech recognition paper)
6. Neural Machine Translation by Jointly Learning to Align and Translate (https://arxiv.org/pdf/1409.0473.pdf) (original seq2seq+attention paper)
7. Attention and Augmented Recurrent Neural Networks (https://distill.pub/2016/augmented-rnns/) (blog post overview)
8. Massive Exploration of Neural Machine Translation Architectures (https://arxiv.org/pdf/1703.03906.pdf) (practical advice for hyperparameter choices)
9. Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models (https://arxiv.org/abs/1604.00788.pdf)
10. Revisiting Character-Based Neural Machine Translation with Capacity and Compression (https://arxiv.org/pdf/1808.09943.pdf)

Events:

Deadlines:

Date: Tue Jan 31

Week 4

Description: Final Projects: Custom and Default; Practical Tips
[slides (slides/cs224n-2023-lecture07-final-project.pdf)] [notes (readings/cs224n-2019-notes06-NMT_seq2seq_attention.pdf)]

Suggested Readings:
1. Practical Methodology (https://www.deeplearningbook.org/contents/guidelines.html) (*Deep Learning* book chapter)

Events: Assignment 4 **out**
[code (assignments/a4.zip)]
[handout (assignments/a4.pdf)]
[latex template (assignments/a4_latex.zip)]
[colab (https://colab.research.google.com/drive/1SMqKVBXkPyqquhQLch_-Pb-FPA9W2scG?usp=sharing)]

Deadlines: Assignment 3 **due**

Date: Thu Feb 2

Description: Self-Attention and Transformers *(by John Hewitt)*
[slides (slides/cs224n-2023-lecture08-transformers.pdf)] [notes (readings/cs224n-self-attention-transformers-2023_draft.pdf)]

Suggested Readings:
1. Default Project Handout (http://web.stanford.edu/class/cs224n/project/default-final-project-bert-handout.pdf)
2. Attention Is All You Need (https://arxiv.org/abs/1706.03762.pdf)
3. The Illustrated Transformer (https://jalammar.github.io/illustrated-transformer/)
4. Transformer (Google AI blog post) (https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html)
5. Layer Normalization (https://arxiv.org/pdf/1607.06450.pdf)
6. Image Transformer (https://arxiv.org/pdf/1802.05751.pdf)
7. Music Transformer: Generating music with long-term structure (https://arxiv.org/pdf/1809.04281.pdf)

Events: Project Proposal **out**
[instructions (project/project-proposal-instructions-2023.pdf)]

Default Final Project **out**
[handout (project/default-final-project-bert-handout.pdf)]

Deadlines:

Date: Tue Feb 7

Week 5

Description: Pretraining *(by John Hewitt)*
[slides (slides/cs224n-2023-lecture9-pretraining.pdf)]

Suggested Readings:
1. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (https://arxiv.org/pdf/1810.04805.pdf)
2. Contextual Word Representations: A Contextual Introduction (https://arxiv.org/abs/1902.06006.pdf)
3. The Illustrated BERT, ELMo, and co. (http://jalammar.github.io/illustrated-bert/)
4. Martin & Jurafsky Chapter on Transfer Learning (https://web.stanford.edu/~jurafsky/slp3/11.pdf)

Events:

Deadlines:

Date: Thu Feb 9

Description: Natural Language Generation *(by Xiang Lisa Li)*
[slides (slides/cs224n-2023-lecture10-nlg.pdf)]

Suggested Readings:
1. The Curious Case of Neural Text Degeneration (https://arxiv.org/abs/1904.09751.pdf)
2. Get To The Point: Summarization with Pointer-Generator Networks (https://arxiv.org/abs/1704.04368.pdf)
3. Hierarchical Neural Story Generation (https://arxiv.org/abs/1805.04833.pdf)
4. How NOT To Evaluate Your Dialogue System (https://arxiv.org/abs/1603.08023.pdf)

Events: Assignment 5 **out**
[code (assignments/a5.zip)]
[handout (assignments/a5.pdf)]
[latex template (assignments/a5_latex.zip)] [colab (https://colab.research.google.com/drive/1VfdgyVFhcPG3ESWwdOCdOllEHnPliO3c?usp=sharing)]

Deadlines: Assignment 4 **due**

Date: Fri Feb 10

Description: Hugging Face Transformers Tutorial Session

3:30 PM - 4:20 PM
Gates B01

Events: Colab (https://colab.research.google.com/drive/1pxc-ehTtnVM72-NViET_D2ZqOlpOi2LH?usp=sharing)

Deadlines:

Date: Tue Feb 14

Week 6

Description: Prompting, Reinforcement Learning from Human Feedback *(by Jesse Mu)*
[slides (slides/cs224n-2023-lecture11-prompting-rlhf.pdf)]

Suggested Readings:
1. Language Models are Few-Shot Learners (https://arxiv.org/abs/2005.14165)
2. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models (https://arxiv.org/abs/2201.11903)
3. Finetuned Language Models Are Zero-Shot Learners (https://arxiv.org/abs/2109.01652)
4. Learning to summarize from human feedback (https://arxiv.org/abs/2009.01325)

Events:

Deadlines: Project Proposal **due**

Date: Thu Feb 16

Description: Question Answering
[slides (slides/cs224n-2023-lecture12-QA.pdf)]

Suggested readings:
1. SQuAD: 100,000+ Questions for Machine Comprehension of Text (https://arxiv.org/pdf/1606.05250.pdf)
2. Bidirectional Attention Flow for Machine Comprehension (https://arxiv.org/pdf/1611.01603.pdf)
3. Reading Wikipedia to Answer Open-Domain Questions (https://arxiv.org/pdf/1704.00051.pdf)
4. Latent Retrieval for Weakly Supervised Open Domain Question Answering (https://arxiv.org/pdf/1906.00300.pdf)
5. Dense Passage Retrieval for Open-Domain Question Answering (https://arxiv.org/pdf/2004.04906.pdf)
6. Learning Dense Representations of Phrases at Scale (https://arxiv.org/pdf/2012.12624.pdf)

Events: Project Milestone **out**
[Instructions (project/CS224N_Final_Project_Milestone_Instructions.pdf)]

Deadlines:

Date: Sat Feb 18

Description:

Events:

Deadlines: Assignment 5 **due**
(11:59 PM)

Date: Tue Feb 21

Week 7

Description: ConvNets, Tree Recursive Neural Networks and Constituency Parsing
[slides (slides/cs224n-2023-lecture13-CNN-TreeRNN.pdf)]

Suggested readings:

1. Convolutional Neural Networks for Sentence Classification (https://arxiv.org/abs/1408.5882.pdf)
2. Improving neural networks by preventing co-adaptation of feature detectors (https://arxiv.org/abs/1207.0580)
3. A Convolutional Neural Network for Modelling Sentences (https://arxiv.org/pdf/1404.2188.pdf)
4. Parsing with Compositional Vector Grammars. (http://www.aclweb.org/anthology/P13-1045)
5. Constituency Parsing with a Self-Attentive Encoder (https://arxiv.org/pdf/1805.01052.pdf)

Events:

Date: Thu Feb 23

Description: Insights between NLP and Linguistics *(by Isabel Papadimitriou)*
[slides (slides/cs224n-2023-lecture14-insights-linguistics.pdf)]

Events:

Date: Tue Feb 28

Week 8

Description: Code Generation *(by Gabriel Poesia)*
[slides (slides/cs224n-2023-lecture15-code-generation.pdf)]

Suggested readings:

1. Program Synthesis with Large Language Models (https://arxiv.org/pdf/2108.07732.pdf)
2. Competition-level code generation with AlphaCode (https://www.science.org/doi/full/10.1126/science.abq1158)
3. Evaluating Large Language Models Trained on Code (https://arxiv.org/abs/2107.03374)

Events:

Deadlines:

Date: Wed Mar 1

Description: Training Large Language Models *(by John Hewitt)*

3:30pm - 4:20pm
Skilling Auditorium

Events:

Deadlines:

Date: Thu Mar 2

Description: Multimodal Deep Learning *(by Douwe Kiela)*
[slides (slides/Multimodal-Deep-Learning-CS224n-Kiela.pdf)]

Events:

Deadlines:

Date: Fri Mar 3

Description:

Events:

Deadlines: Project Milestone **due**

Date: Tue Mar 7

Week 9

Description: Coreference Resolution
[slides (slides/cs224n-2023-lecture17-coref.pdf)]

Suggested readings:
1. Coreference Resolution Chapter from Jurafsky and Martin (https://web.stanford.edu/~jurafsky/slp3/21.pdf)
2. End-to-end Neural Coreference Resolution (https://arxiv.org/pdf/1707.07045.pdf)

Events:

Deadlines:

Date: Thu Mar 9

Description: Analysis and Interpretability Basics *(by John Hewitt)*
[slides (slides/cs224n-2023-lecture18-analysis.pdf)]

Events:

Deadlines:

Date: Fri Mar 10

Description: Latex Tutorial *(by Rishi Desai)*

3:30pm - 4:20pm
Skilling Auditorium

Events:

Deadlines:

Date: Tue Mar 14

Week 10

Description: Model Interpretability and Editing *(by Been Kim)*
[slides (slides/Been-Kim-StanfordLectureMarch2023.pdf)]

Events:

Deadlines:

Date: Thu Mar 16
Description: Final Project Emergency Assistance (no lecture)
Extra project office hours available during usual lecture time, see Ed.
Events:
Deadlines:

Date: Sat Mar 18
Description:
Events:
Deadlines: Project **due** [instructions (project/final-report-instructions-2023.pdf)]

Date: Monday Mar 20
Description: Poster Session
5pm-9pm [More details (project.html)] Location: Tressider Oak Lounge
Events:
Deadlines: [Printing guide (project/poster-printing-guidelines-2023.pdf)]