CS371N Midterm Topics

The exam will have the same broad types of questions (multiple choice, short answer, long answer) as past midterms. For long answer questions, you can show your work for partial credit, but for our sake and yours, please try to refrain from writing an essay as a response unless the question explicitly asks for it.

Topics   Below is the list of topics that will be covered on the midterm, which is the bulk of the course material thus far.

- Bag-of-words features: how these feature spaces look and how they work for classification

- Perceptron (binary): algorithm, loss function

- Logistic regression (binary): model (the LR formula), training objective, gradient update

- Sentiment analysis: what kinds of features are useful?

- Multiclass classification: how weights and features work in this setting

- Multiclass perceptron: model definition, how to train it

- Optimization: stochastic gradient descent, impact of step size on optimization, impact of initialization

- Feedforward neural networks: definition, initialization

- Training neural networks

- Word embeddings: skip-gram model, definition, properties, potential for bias

- Deep averaging networks: model from Assignment 2, limitations of the model

- n-gram language modeling: basic definition, count-based parameter estimation

- Self-attention: mathematical definition, how it works

- Transformers: architecture details

- Transformer Language Modeling: everything from Assignment 3, including training, inference, and evaluation (perplexity)

- Encoders: masked language modeling and BERT

- Subword tokenization

- Inference in decoders, nucleus sampling, beam search

- POS tagging: understanding ambiguities (like the *Fed raises...* example)

- Hidden Markov Models: definition, parameter estimation (count and normalize), Viterbi algorithm

- Constituency syntax: what trees look like, ambiguities (*Ban on nude dancing...*, *eat spaghetti with chopsticks*)

- PCFGs: definition, parameter estimation (counting and normalizing), what binarization and Markovization are, CKY algorithm

- **Not covered:** dependency parsing, as it's so close to the midterm

Other content   You should expect to see examples of text and be comfortable reasoning about how these algorithms might work on such examples, as in the assignments so far. We won't expect you to know things like part-of-speech definitions or have encyclopedic knowledge of grammar structures—we will provide the necessary information for such questions.

Readings   We won't expect you to know content from the Eisenstein book or other readings that hasn't been covered in lectures or assignments.

Practice problems   The best source of practice problems is the prior midterms. Language modeling and Transformers are a new topic now. Questions for these will largely revolve around content in the lectures and assignments.

Note that parsing and tagging have been a bit de-emphasized, so the midterm will not focus as heavily on structured prediction questions (especially Viterbi) as past years' midterms have.