CS378 Spring 2020 Midterm Topics

For the midterm, **you will be allowed one double-sided 8.5x11 note sheet (typed is okay)**. We may also choose to print critical formulas on the exam if these are necessary for a question. Below is the list of topics that will be covered on the midterm, which is the bulk of the course material up to and including March 5.

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

- Perceptron (binary): algorithm, loss function

- Logistic regression (binary): model, training objective

- Sentiment analysis

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

- Multiclass perceptron: definition and training

- Optimization: stochastic gradient descent, step sizes, initialization

- Feedforward neural networks

- Training neural networks

- Word embeddings: skip-gram, skip-gram with negative sampling

- POS tagging

- Hidden Markov Models: definition, parameter estimation, Viterbi algorithm

- Beam search

- Conditional random fields: intuition and general concepts (you won't be expected to know formulas)

- PCFGs: definition, parameter estimation, grammar refinement, CKY algorithm

- Shift-reduce dependency parsing (arc-standard system)

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 that hasn't been covered in lecture, even if it's in the posted readings. As for the other assigned readings, we may ask questions pertaining to concepts from these (concepts being at the level discussed in lecture), but we generally won't assume that you've committed the specific approaches to memory.

Practice problems   The best source of practice problems is last year's midterm, which is posted on the course website. Note that Naive Bayes classification (multiple choice Q1-4) was dropped this year.

The book's problems are more theoretical and challenging than we would have for an exam. The following book exercises are reasonable review problems that cover concepts from lecture: Chapter 3 Ex 2, Chapter 7 Ex 1, Chapter 9 Exs 8 and 9, Chapter 10 Ex 4, Chapter 11 Ex 1, Chapter 14 Exs 4 and 5.