

CS 4650 and 7650 Midterm Review

Jacob Eisenstein

The goal of the midterm is to evaluate your understanding of the core concepts in this course: algorithms, representations, and challenges for natural language processing. It will emphasize areas that were not covered by the projects and homeworks, although it may include anything from the course lectures or reading. **The exam is closed-book and closed-laptop, but you may bring a single two-sided sheet of notes.**

Here are some key topics in the course so far:

- Linguistic fundamentals
 - Word sense ambiguity
 - Morphology
 - Parts-of-speech
 - Context-free grammars and constituents
 - Dependency grammar
- Statistical NLP
 - The bag-of-words feature function representation
 - Naive Bayes and generative models
 - Perceptron and logistic regression
 - Structured prediction (HMM, structured perceptron, and CRF)
 - Expectation maximization
 - Relative frequency estimation and smoothing
- Algorithms and representations for NLP
 - Finite-state automata and transduction
 - Viterbi, forward, and backward
 - CKY and Inside algorithms
 - Regular and context-free languages
 - N-gram language models

The exam will generally avoid lengthy derivations or linguistic analyses that require deep understanding of English, and will focus on “thought questions” about the core concepts. However, some derivation and analysis may be required. In the past, midterm performance has been strongly correlated with doing the course reading.