6.891 (Fall 2003): Machine Learning Approaches for Natural Language Processing

Michael Collins
mcollins@ai.mit.edu

Course Coverage

- **NLP sub-problems:** part-of-speech tagging, parsing, word-sense disambiguation, etc.
- **Machine learning techniques:** probabilistic context-free grammars, hidden markov models, estimation/smoothing techniques, the EM algorithm, log-linear models, cotraining, etc.
- **Applications:** information extraction, machine translation, natural language interfaces...

Evaluation

- Survey paper (30% of final grade). Read three papers on a particular research project, and write a paper that synthesizes and compares the results of the three papers.
- Project (70% of final grade). A course project on some research problem in machine learning for natural language.

Intended goals of the course

- By the end of the course, you should have the background to read research papers in the general area of statistical NLP; and be able to start research in the area.
- Techniques and problems should be of interest to researchers in related fields: e.g., speech recognition, computational biology, linguistics, brain and cognitive sciences (psycholinguistics), machine learning, statistics, computer vision, probabilistic methods in AI.

Background required for the course

- Course should be fairly self-contained; **but** a background in basic probability theory will be very helpful.

Readings for the course

- I'll give pointers to relevant papers as we go along.
- No set textbook, but *Manning and Schutze: Foundations of Statistical Natural Language Processing; MIT Press* should be useful.

Syllabus

- Introduction (1 lecture)
- Estimation techniques, and language modeling (1 lecture)
- Stochastic parsing (3 lectures)
- Stochastic tagging, and maximum entropy/log-linear models (2 lectures)
- The EM algorithm for PCFGs, HMMs, and machine translation (2 lectures)
- Information extraction (2 lectures)
- Machine translation (3 lectures)
- Vision and language (1 lecture)
- Dialogue systems (2 lectures)
- NLP issues in different languages,
  e.g., Czech, Chinese, Arabic, German (1 lecture)
- Word-sense disambiguation (1 lecture)
- Semi-supervised approaches (2 lectures)
- Discriminative/reranking approaches, and kernels over NLP structures (2 lectures)
- Learning in Optimality Theory (1 lecture)
- Conclusions/open problems (1 lecture)