Provable Bounds in Machine Learning

August 1st

Speaker: Michael Collins (10-11)

Title: Provable ML Methods for Natural Language Processing

Abstract: In this talk I'll attempt to give a broad overview of machine learning problems for NLP, with an emphasis on provable methods. I'll focus on two topics:

1) Models with latent variables. Latent-variable models have had tremendous impact on NLP. The EM algorithm is widely used for parameter estimation with these models, but in general has no guarantee of reaching the global maximum of the likelihood function. I'll first describe some of our recent work on spectral learning of probabilistic context-free grammars with latent variables. I'll then describe open problems in areas such as statistical machine translation, and speech recognition.

2) Linear programming relaxations, and Lagrangian relaxation, for inference in NLP. Many inference problems in NLP can be framed as combinatorial optimization problems, where the goal is to find the highest scoring "structure" out of a large set of possible structures (for example, the highest scoring translation for a given input sentence). These problems are often challenging: either polynomial time but with impractically high cost, or NP hard. I'll describe recent work on Lagrangian relaxation methods for a number of problems in NLP. The good news is that these methods often produce exact solutions, with certificates of optimality; the bad news is that at present there is little theoretical understanding of why that is the case.

The first part of the talk includes joint work with Shay Cohen, Karl Stratos, Dean Foster, and Lyle Ungar. The second part of the talk covers joint work with Yin-Wen Chang, Tommi Jaakkola, Terry Koo, Sasha Rush, and David Sontag.

Speaker: Sanjeev Arora (11-12)

Title: Is Learning Easy? Three vignettes.

Abstract:
The way machine learning is formalized often leads to computationally intractable problems. I will talk about three recent results that suggest how to get around this barrier of intractability.
Vignette 1: What binary classifiers should one learn?
Vignette 2: Learning topic models: Is there life beyond max-likelihood?

Vignette 3: Can we have provable algorithms for deep learning?

Speaker: Sham Kakade (1:30-2:30)

Title: Scalable Spectral Methods for Learning Latent Variable Models

Speaker: Percy Liang (2:30-3:30)

Title: Learning Latent-Variable Models of Natural Language

Abstract: A key property of natural language is that raw observations (e.g., sentences) are often associated with latent structures (e.g., parse trees). To infer these latent structures, we need to design sensible probabilistic models that connect latent structures with observations, as well as develop efficient algorithms for estimating the model parameters. First, I will discuss syntactic and semantic parsing models, showing how the latter can be used for question answering. Second, I will present recent work on learning restricted PCFGs using eigenvalue methods.

Speaker: David Sontag (4-5)

Title: Probabilistic Inference as Statistical Recovery

Abstract: In recent years, the machine learning community has had substantial empirical success with using linear programming relaxations for probabilistic inference in graphical models with complex interactions between the variables. In this talk, I will interpret these applications of probabilistic inference as that of statistical recovery. I will argue that these empirical successes should not have been surprising because the corresponding inference problems were operating in an information rich regime where inference is not fundamentally difficult. I will then discuss the information poor regime, where I believe there is a substantial need for new approaches for marginal inference, and the intermediate regime, where (in theory) hard combinatorial optimization problems seem to arise.

Speaker: Yann LeCun (5-6)

Title: Deep Learning: the Theoretician's Nightmare or Paradise?

August 2nd

Speaker: John Lafferty (10-11)

Title: Dictionary Learning and Sparsified Covariance Matrices for Linear Estimation

Speaker: Rob Schapire (11-12)

Title: Understanding AdaBoost

Abstract: Boosting is an approach to machine learning based on the idea of creating a highly accurate prediction rule by combining many relatively weak and inaccurate rules. AdaBoost, the first practical boosting algorithm, has enjoyed empirical success in a number of fields, and a remarkably rich theory has evolved to try to understand how and why it works, and under what conditions. At various times in its history, AdaBoost has been the subject of controversy for the mystery and paradox it seems to present with regard to this question. This talk will give a high-level review and comparison of the varied attempts that have been made to understand and "explain" AdaBoost. These approaches (time permitting) will include: direct application of the classic theory of Vapnik and Chervonenkis; the margins theory; AdaBoost as a loss-minimization algorithm (possibly implicitly regularized); and AdaBoost as a universally consistent method. Both strengths and weaknesses of each of these will be discussed.

Speaker: Nina Balcan (1:30-2:30)

Title: Incorporating Unlabeled Data and Interaction in the Learning Process

Abstract: Many modern machine learning applications -- ranging from spam detection to computational biology to computer vision -- have huge quantities of raw data available, much more than can actually be labeled for training.

In order to better use all this data a number of powerful new learning approaches have been proposed and explored. For settings in which labeling can be done interactively, a powerful technique is Active Learning, in which the algorithm adaptively chooses informative examples to be labeled in order to dramatically (and sometimes exponentially) reduce labeling effort. When labeling cannot be done interactively and the labeled sample is given a priori, Semi-Supervised learning uses both the labeled and (a lot more) unlabeled data to achieve better performance than learning over just labeled data alone. In this talk we discuss new theoretical frameworks as well as new algorithms for Active Learning and Semi-Supervised Learning that both explain and exploit the benefit of unlabeled data and interaction for learning.

Speaker: Alexander Rakhlin (3-4)

Title: Online Learning: A Minimax Analysis and an Algorithmic Framework

Abstract: We give an overview of various directions in the field of online learning. We present a minimax framework and then show a principled way of deriving online learning algorithms from the minimax analysis. Various known methods and new algorithms are shown to seamlessly follow from this approach. Our

framework also captures such ``unorthodox'' methods as Follow the Perturbed Leader and the R^2 forecaster. We present a family of randomized methods that use the idea of a ``random playout''. Finally, we discuss future research directions.