

Introduction to Classification: Likelihoods, Margins, Features, and Kernels

Tutorial for NAACL-HLT 2007

Dan Klein
Computer Science Division
University of California, Berkeley
klein@cs.berkeley.edu

Overview

Statistical methods in NLP have exploited a variety of classification techniques as core building blocks for complex models and pipelines. In this tutorial, we will survey the basic techniques behind classification. We first consider the basic principles, including the principles of maximum likelihood and maximum margin. We then discuss several core classification technologies: naive Bayes, perceptrons, logistic regression, and support vector machines. The discussion will include the key optimization ideas behind their training and the empirical trade-offs between the various classifiers. Finally, we consider the extension to kernels and kernelized classification: what can kernels offer and what is their cost? The presentation is targeted to NLP researchers new to these methods or those wanting to understand more about how these techniques are interconnected.

Topics

1. Basics of classification
 - (a) Feature-based representations
 - (b) Linear classifiers
 - (c) Principles of classification: likelihood and margin
 - (d) Smoothing and regularization
 - (e) Structured classification
2. Specific techniques
 - (a) Perceptrons
 - (b) Naive Bayes
 - (c) Logistic regression / maximum entropy
 - (d) Support vector machines
 - (e) Comparison and trade-offs
3. Kernel methods
 - (a) Why kernels (and why not)?
 - (b) Kernelized linear classifiers
 - (c) Kernelized perceptrons
 - (d) Kernelizing SVMs and logistic regression
 - (e) Advanced kernels and structure