Final Exam Study Topics
Machine Learning 10-701
Spring 2011

Tom Mitchell

Looking back on the course, here are some of the key concepts we have covered. Check your understanding of these topics as you study for the final exam.

What is learning:
- improving performance P at task T through experience E
- we often formulate learning problems as estimating some parameters to optimize some objective

Objective functions to optimize when learning parameters
- maximize likelihood
- maximize conditional likelihood (e.g., Logistic regression)
- maximize a posteriori estimates (MAP) and the use of priors
- maximize margin (e.g., for SVM's)
- minimize sum of squared errors (e.g., in linear regression)

Overfitting, what it is, how to detect it, how to avoid it
- cross validation
- role of priors and MAP estimates
- PAC theory as a way of characterizing, bounding overfitting

Bayesian networks as a general language for probabilistic models
- representation (conditional independence, D-separation, factoring the joint distribution)
- inference
- learning

Classification:
- many algs.: decision trees, naïve Bayes, logistic regression, SVM, neural network, weighted majority, Fisher linear discriminant, …
- be able to derive your own, by choosing a representation, an objective, and deriving a training method to optimize it
- gradient descent as a general weak method for optimization
- generative and discriminative classifiers (P(X,Y) versus P(Y|X))

Regression:
- probabilistic model where MLE = minimize sum of squared errors
- MAP estimates and regularization
- kernel regression

Semi-supervised learning - learning when some variables are unobserved
- Expectation Maximization (EM)
- Hidden Markov Models
- Cotraining

Learning representations - projecting data into new spaces
- Neural networks, hidden units
- Principle Components Analysis (PCA), minimizing reconstruction error
- many others: CCA, ICA, Fisher discriminant, topic models (Latent Dir. Alloc.)
- supervised and unsupervised learning of representations

Kernel methods
- kernel linear regression
- using kernels to operate virtually in a higher dimensional space
- primal and dual forms of the optimization problem

Maximum margin classification
- hard margin SVM's
- soft margin SVM's
- use of kernels in SVM's
- PAC results for SVM's

Active learning
- pool based active learning
- uncertainty sampling
- query by committee

Markov decision processes
- the probabilistic model
- temporal difference learning, Q learning
- convergence guarantees