Midterm Review

Machine Learning 10-601

Tom M. Mitchell
Machine Learning Department
Carnegie Mellon University

October 25, 2011

See practice exams on our website

Midterm is in class October 27

Midterm is open book, open notes, NO computers, NO internet

Covers all material presented up through today’s class.

Some Topics We’ve Covered

**Decision trees**
- entropy, mutual info., overfitting

**Probability basics**
- Bayes rule, MLE, MAP, conditional indep.

**Naïve Bayes**
- conditional independence,
- # of parameters to estimate, decision surface

**Logistic regression**
- form of P(Y|X)
- generative vs. discriminative

**Linear Regression**
- minimizing sum sq. error (why?)
- regularization ~ MAP

**Sources of Error**
- unavoidable error, bias, variance

**Overfitting, and Avoiding it**
- priors over H
- cross validation
- PAC theory: probabilistic bound on overfitting

**Bayesian Networks**
- factored representation of joint distribution, conditional independence assumptions, D-separation
- inference in Bayes nets
- learning from fully/partly observed data

**PAC Learning**
- sample complexity
- probabilistic bounds on error_{train} – error_{true}
- VC dimension
Understanding/Comparing Learning Methods

Naïve Bayes

Form of learned model
• Inputs:
• Outputs:

Optimization Objective:

Algorithm:

Assumptions:

Guarantees?:

Decision boundary:

Generative/Discriminative?

Logistic Regression

Form of learned model
• Inputs:
• Outputs:

Optimization Objective:

Algorithm:

Assumptions:

Guarantees?:

Decision boundary:

Generative/Discriminative?
Four Fundamentals for ML

1. Learning is an optimization problem
   - many algorithms are best understood as optimization algs
   - what objective do they optimize, and how? Local minima?
   - gradient descent/ascent as general fallback approach

2. Learning is a parameter estimation problem
   - the more training data, the more accurate the estimates
   - MLE, MAP, M(Conditional)LE, …
   - to measure accuracy of learned model, we must use test (not train) data
Four Fundamentals for ML

1. Learning is an optimization problem
   - many algorithms are best understood as optimization algs
   - what objective do they optimize, and how?
2. Learning is a parameter estimation problem
   - the more training data, the more accurate the estimates
   - MLE, MAP, M(Conditional)LE, …
   - to measure accuracy of learned model, we must use test (not train) data
3. Error arises from three sources
   - unavoidable error, bias, variance
   - PAC learning theory: probabilistic bound on overfitting: $\text{error}_{\text{true}} - \text{error}_{\text{train}}$

Bias and Variance of Estimators

given some estimator $Y$ for some parameter $\theta$, we note $Y$ is a random variable (why?)

the bias of estimator $Y$: $E[Y] - \theta$
the variance of estimator $Y$: $E[(Y - E[Y])^2]$

consider when
- $\theta$ is the probability of “heads” for my coin
- $Y =$ proportion of heads observed from 3 flips
consider when
- $\theta$ is the vector of correct parameters for learner
- $Y =$ parameters output by learning algorithm
Four Fundamentals for ML

1. Learning is an optimization problem
   - many algorithms are best understood as optimization algs
   - what objective do they optimize, and how?

2. Learning is a parameter estimation problem
   - the more training data, the more accurate the estimates
   - MLE, MAP, M(Conditional)LE, …
   - to measure accuracy of learned model, we must use test (not train) data

3. Error arises from three sources
   - unavoidable error, bias, variance
   - PAC learning theory: probabilistic bound on overfitting: \( \text{error}_{\text{true}} - \text{error}_{\text{train}} \)

4. Practical learning requires making assumptions
   - Why?
   - form of the \( f:X \rightarrow Y \), or \( P(Y|X) \) to be learned
   - priors on parameters: MAP, regularization
   - Conditional independence: Naive Bayes, Bayes nets, HMM’s