Midterm Review

Machine Learning 10-701

Tom M. Mitchell
Machine Learning Department
Carnegie Mellon University

March 1, 2011

See practice exams on our website
Attend recitation tomorrow

Midterm is open book, open notes, NO computers

Covers all material presented up through today's class.

Some Topics We've Covered

Decision trees
- entropy, mutual info., overfitting

Probability basics
- rv's, manipulating probabilities,
- 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

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

Clustering
- mixture of Gaussians, EM

Understanding/Comparing Learning Methods

Form of learned model

- Inputs:
- Outputs:

Optimization Objective:

Algorithm:

Assumptions:

Guarantees?:

Decision boundary:

Generative/Discriminative?

	Naïve Bayes	Logistic Regression
Form of learned model	want $P(Y\|X) \leftarrow \frac{P(X\|Y)P(Y)}{P(X)}$	$\boxed{P(Y\|X)} \leftarrow$ directly
• Inputs:	$\langle X_1, \dots X_n \rangle$ real or discrete vals	"
• Outputs:	discrete valued Y	" $\quad \frac{\exp(\sum w_i x_i)}{1+\exp(\;)}$
Optimization Objective:	MLE, MAP	M(Conditional)LE
Algorithm:		
Assumptions:	cond indep $\leftarrow$ repr $\leftarrow$ learning	
Guarantees?:		learning makes no explicit cond indep assump.
Decision boundary:		
Generative/Discriminative?		

Four Fundamentals for ML

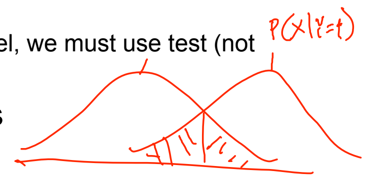
1. Learning is an optimization problem
   - many algorithms are best understood as optimization algs
   - what objective do they optimize, and how?

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

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



Bias and Variance

given some estimator Y for some parameter θ, 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

- θ is the probability of “heads” for my coin
- Y = proportion of heads observed from 3 flips

consider when

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