Midterm Review

Machine Learning 10-701

Tom M. Mitchell
Machine Learning Department
Carnegie Mellon University

March 1, 2010

See practice exams at:
http://www.cs.cmu.edu/~tom/10601_sp09/601-sp09-midterm-solutions.pdf
http://select.cs.cmu.edu/class/10701-F09/exams.html

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, overfitting

Probability basics

rv's, manipulating probabilities,
Bayes rule, MLE, MAP,
conditional indep.

Instance-based learning

nearest nbr., density estimation,
Bayes optimal classifier

Naïve Bayes

conditional indep, # of parameters
to estimate,

Logistic regression

form of P(Y|X) implied by N. Bayes,
generative vs. discriminative

Linear Regression

minimizing sum sq. error ~ MLE
regularization ~ MAP, non-linear

Neural Networks

gradient descent,
learning hidden representations

Model Selection

overfitting, bias-variance

Clustering

k-means, mixture Gaussians, EM

Hidden Markov Models

time series model, backward-forward

Bayesian Networks

factored representation of joint
distribution, encoding conditional
independence assumptions

	representation of P(Y|X)	decision surface	optimization objective	convergence guarantee?	other assumptions?
Naïve Bayes					
Logistic Regr.					
Linear Regr.					
Neural net					
Dec. Tree					
Gaussian Mixture model					
HMM					
Bayes Net					
kNN					

Four Fundamentals for ML

1. Learning is an optimization problem

2. Learning is a parameter estimation problem

3. Error arises from three sources

4. Practical learning requires modeling assumptions, such as …

Learning is an optimization problem

- many algorithms are best understood as optimization algs
- what objective do they optimize, and how?
- naïve Bayes? logistic regression? linear regression?

suppose want $f(x) = Y$, $P(Y|X)$

Learning is parameter estimation

θ defines learned f.

HMM

$Y_1 \rightarrow Y_2 \rightarrow Y_3$, $\downarrow$ X_1 X_2 X_3

- the more training data, the more accurate the estimates
- to measure accuracy of learned model, we must use test (not train) data
- cross validation

$P(data|\theta) = \prod_k P(X_1^k \ldots X_3^k | \theta)$

N Bayes $P(Y|X_1 \ldots X_n) = \prod_i P(X_i|Y) \cdot P(Y)$

$\hat{P}(Y|X, \theta)$ ← params

training examps: $X^1 Y^1, X^2 Y^2 \ldots X^K Y^K$

$P(data|\theta) =$ data likelihood $(\theta) = \prod_{k=1}^{K} P(X^k Y^k | \theta)$

MCLE → cond. likelihood $= \prod_k P(Y^k | X^k, \theta)$

MLE = $\arg\max_\theta$ data likelihood $(\theta) = \prod_{k=1}^{K} P(X^k | Y^k, \theta) \, P(Y^k|\theta)$

MAP ≠ $\arg\max_\theta P(\theta | data) = \dfrac{P(data|\theta) \, P(\theta)}{P(data)}$

if assume $Y = f(X, \theta) + \epsilon$, $\epsilon \sim N(0, \sigma)$

Lin regr. $\hat{\theta} = \arg\max_\theta \prod_k P(Y^k | X^k, \theta)$

$\hat{\theta} = \arg\min_\theta \sum_k (Y^k - \hat{f}(X^k, \theta))^2$

Error arises from three sources

– Bayes optimal error, bias², variance

Bayes optimal error: inescapable error even if know perfectly $P(Y|X)$

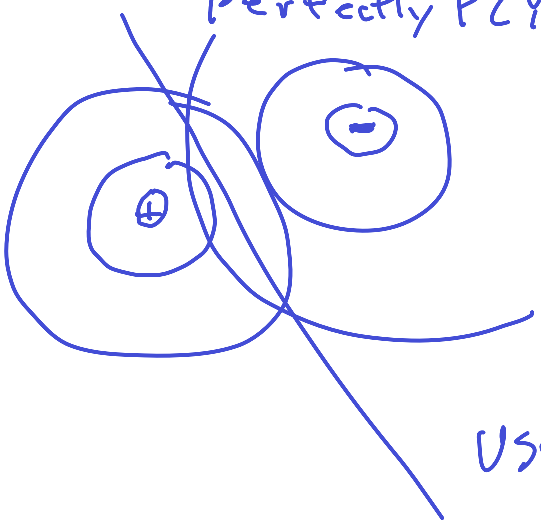
bias: $E[\hat{\theta}] \neq \theta$

$E[\hat{\theta}] - \theta$

variance: $E[\hat{\theta} - E[\hat{\theta}]]$

learning $f_\theta(x)$ means picking θ

r.v. because sample of training data is drawn randomly from $P(X,Y)$



Usually simpler models exhibit higher bias, lower variance

more data → lower variance

little data

3 students/classroom

lots

3000000s/cl

Bias and Variance

$= \hat{\theta}$ estimate

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

the bias of estimator Y : $E[Y] - \theta$ — if Y is unbiased then $E[Y] = \theta$

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

expectation is over different draws of training data

consider when

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

Practical learning requires making assumptions

- Why?
- form of the f:X → Y, or P(Y|X), or P(…) to be learned
- priors on parameters → MAP, regularization
- Conditional independence → Naive Bayes, Bayes nets

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
   - Bayes optimal error, bias, variance
4. Practical learning requires modeling 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