

Lecture 26: Semester review

CSI 5v93: Introduction to machine learning

Baylor University
Computer Science Department

Dr. Greg Hamerly
<http://cs.baylor.edu/~hamerly/>

CSI 5v93: Introduction to machine learning, Lecture 26 – p. 1/9

Questions?

CSI 5v93: Introduction to machine learning, Lecture 26 – p. 2/9

Semester topics:

- supervised learning
- linear regression
- linear classification
- naive Bayes learning
- support vector machines
- unsupervised learning

Supervised learning

- inputs and outputs
- bias/variance tradeoff
- overfitting and underfitting
- statistics: expected values
- curse of dimensionality
- learning of parameters
- Bayes' rule
- maximum likelihood estimation
- log-likelihood
- complexity restrictions
- linear model
- k -nearest neighbor model
- cross-validation
- confusion matrix

Linear regression

- linear model
- basis expansions
- least squares criterion and its solution
- variable selection
 - hypothesis testing
 - z test
 - F test
 - subset selection
 - stepwise selection
 - coefficient shrinkage (ridge regression)

Linear classification

- linear boundaries for classification
- discriminant functions $\delta(x)$ and decision boundaries
- linear regression with multiple classes
 - indicator matrix
 - masking
 - learning the models
- posterior probabilities $P(\text{class} = c | X = x)$
- log-odds
- multivariate Gaussians
- linear discriminant analysis (LDA)
- LDA parameter estimation
- quadratic discriminant analysis (QDA)
- regularized discriminant analysis
- logistic regression

Naive Bayes learning

- Bayes' rule and using it for classification
- conditional independence assumption
- ML and MAP classifications
- frequency-based probability model
- prior probabilities
- Laplace smoothing of zero probabilities
- log-probabilities
- binning continuous variables (equal-width, equal-frequency)
- shaping probabilities
- ranking probabilities
- two probability models for text documents:
 - model each word as occurring/not occurring in a document – multivariate Bernoulli model
 - model each word that occurs (counting multiple occurrences) – multinomial model
- ranking features with conditional entropy

CSI 5v93: Introduction to machine learning, Lecture 26 – p. 7/9

Support vector machines

- separating hyperplanes
- maximizing the margin
- support vectors
- “slack” variables for non-separable data
- basis expansions
- kernel functions and the kernel trick
- mathematical definitions of SVM

CSI 5v93: Introduction to machine learning, Lecture 26 – p. 8/9

Unsupervised learning

- learning without a training signal
- goals and motivations of unsupervised learning
- types of unsupervised learning
- data clustering algorithms
- k -means algorithm
- hierarchical bottom-up clustering