Stony Brook University
CSE 512 - Machine Learning
Tentative Syllabus

Basics

Lecture 1:

- What is ML? ML applications
- Learning paradigms
    - Supervised learning (classification, regression)
    - Unsupervised learning (density estimation, clustering, dimensionality reduction)
- Bayes Optimal Learning Rule

Readings: Bishop 2.1, Appendix B

(Optional) Mitchell, Ch 1

(Optional) Murphy, 1.1, 1.2, 1.3.1

Lecture 2:

- Point estimation
    - Loss functions
    - Maximum Likelihood Estimation (MLE)
    - Maximum A Posterior (MAP) Estimation
    - Bayesian learning
    - MLE vs. MAP
    - Bayes Optimal Classifier

http://brenocon.com/blog/2009/07/beta-conjugate-explorer/

Readings: Bishop: Sec 1.5, 2.2, 2.3 (up to 2.3.6)

(Additional Resource) Andrew Moore's basic probability tutorial

Linear Models (Regression, Classification)

Lecture 3:

- Gaussians,
- Linear Regression, [Applet]
- Overfitting,
- Bias-Variance Tradeoff,
- What's ML revisited

Readings: Bishop 1.1 to 1.4,

Bishop 3.1, 3.1.1, 3.1.4, 3.1.5, 3.2, 3.3, 3.3.1, 3.3.2

(Additional Resource) Andrew Moore's Tutorial on regression

(Recommended) Hastie, Ch 7

(Optional) Murphy, 1.4, Ch 7

Lecture 4:

- Conditional Independence,
- Naive Bayes, [Applet]
- Gaussian Naive Bayes

Readings: Bishop 1.3, 1.5, 3.2,

Mitchell's Chapter on Naive Bayes and Logistic Regression (Sect. 1 and 2)

(Optional) Murphy, 3.4

Lecture 5:

- Generative v. Discriminative [Applet]
- Logistic Regression [Applet]

Readings: Mitchell's Chapter on Naive Bayes and Logistic Regression (Sect. 1 and 2)

Bishop - 4.0, 4.2, 4.3, 4.4, 4.5

(Optional) Murphy, Ch 8

(Optional) Ng and Jordan's NIPS 2001 paper on Discriminative versus Generative Learning [pdf] [ps]

Non-linear models and Model selection

Lecture 6:

- Decision Trees [Applet]
- Entropy, Information Gain
- Overfitting, Pre-and Post-pruning

Readings: (Bishop - 1.6) Information Theory

(Bishop - 14.4) Tree-based Models

(Recommended) Mitchell, Ch 3

(Recommended) Quantities of Information Wikipedia entry

(Recommended) Nils Nilsson's Chapter (All Sections): Decision Trees

(Optional) Murphy, 16.2

Lecture 7:

- Combining weak classifiers
- Adaboost algorithm [Adaboost Applet]
- Comparison with logistic regression and bagging

Readings: (Bishop 14.3) Boosting

Boosting homepage

Schapire Boosting Tutorial and its [Video].

(Optional) Multi-class AdaBoost by Zhu, Rosset, Zou, and Hastie.

(Optional) Murphy, 16.4

Lecture 8:

- Cross Validation,
- Simple Model Selection,
- Regularization,
- Information Criteria (AIC, BIC, MDL)

Readings: (Bishop 1.3) Model Selection / Cross Validation

(Bishop 3.1.4) Regularized least squares

(Hastie et al. 3.2, 3.3, 3.4) Model selection and L1 regularization

(Optional) Ron Kohavi's paper, A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection.

(Additional Resource) Minimum Description Length website

Lecture 9:

- Neural Nets [Applet]
- Prediction – Forward-propagation
- Training – Back-propagation

Readings: (Bishop 5.1) Feed-forward Network Functions

(Bishop 5.2) Network Training

(Bishop 5.3) Error Back-propagation

(Additional Resource) [CMU Course] on Neural Nets

(Optional) Murphy, 16.5

Lecture 10:

- Instance-based Learning (Nonparametric methods) [Applet]
- Histogram, Kernel Density Estimation
- K-NN Classifier
- Kernel Regression

Readings: (Bishop 2.5, 6.3) Nonparametric Methods

(Recommended) Mitchell, Ch 8

Andrew Moore's Tutorial on Instance-based Learning

Margin-based approaches

Lecture 11:

- SVMs [LibSVM Applet] [Another SVM Applet]
- Maximum Margin Classifiers
- Slack variables, Hinge loss

Readings: (Bishop 7.1, 4.1.1, 4.1.2, Appendix E)

Hearst 1998: High Level Presentation

Burges 1998: Detailed Tutorial

(Additional Resource) Smola video tutorial on SVM (see Part 3)

(Additional Resource) Scholkopf video tutorial on kernels

(Additional Resource) http://www.svms.org

(Optional) Murphy, 14.5

Lecture 12:

- SVMs – Duality and The Kernel Trick

Readings: (Bishop 6.1, 6.2) Kernels

(Additional Resource) http://www.kernel-machines.org

(Optional) Murphy, 14.4

Learning Theory

Lecture 13:

- PAC-learning [Applets]
- Sample complexity
- Haussler bound, Hoeffding's bound

Readings: Goldman's COLT survey, sections 1-3.1

(Recommended) Mitchell Ch 7

(Optional) John Langford's tutorial on generalization bounds

(Additional Resource) Langford video tutorial on generalization bounds

(Additional Resource) John Shawe-Taylor video tutorial on statistical learning theory

(Additional Resource) http://www.learningtheory.org

Lecture 14:

- VC Dimension
- Mistake Bounds
- Midterm exam review

Readings: (Recommended) Mitchell Ch 7

(Optional) Littlestone's original (excellent) paper on the Mistake Bound model: Learning Quickly When Irrelevant Attributes Abound: A New Linear-Threshold Algorithm

Structured Models (Graphical Models and HMM)

Lecture 15:

- Bayesian Networks – Representation
  [Applet: Java Bayes] [Another Bayes net applet]
    - Factorization of joint distribution
    - Local Markov Assumption
    - D-separation
    - Representation Theorem

Readings: (Bishop 8.1, 8.2) Bayesian Networks

Intro to Graphical Models by K. Murphy

(Additional resource) Heckerman BN Learning Tutorial

(Optional) Murphy, Ch 10

Lecture 16:

- Bayesian Networks – Inference
    - Marginalization
    - Variable Elimination

Readings: (Bishop 8.4.1, 8.4.2) - Inference in Chain/Tree Structures

Lecture 17:

- Bayesian Networks – Structure Learning
    - Learning CPTs
    - Learning structure - Chow-Liu Algorithm

Readings: (Additional resource) Koller et. al, Graphical Models in a Nutshell

(Optional) Murphy, 26.1-26.4

Lecture 18:

- HMMs
    - Representation
    - Forward Algorithm
    - Forward-Backward Algorithm
    - Viterbi Algorithm
    - Baum-Welch Algorithm

Readings: (Bishop, Ch 13)
HMM and EM Tutorial

(Optional) Rabiner's Detailed HMMs Tutorial

(Additional Resource) Ghahramani, "An introduction to HMMs and Bayesian Networks"

(Optional) Murphy, Ch 17

Unsupervised and semi-supervised learning

Lecture 19:

- Clustering
    - Hierarchical Clustering
    - Spectral Clustering [Spectral Clustering demo]

Readings: Spectral Clustering tutorial by Ulrike von Luxburg

(Optional) Murphy, 25.4, 25.5

Lecture 20:

- Clustering
    - K-Means [Applet: K-means]
    - Gaussian Mixture Model [Applet: Mixture of Gaussians]

Readings: (Bishop 9.1, 9.2) - K-means, Mixtures of Gaussian

Lecture 21:

- Expectation Maximization (EM)

Readings: (Bishop 9.3, 9.4) – EM

Neal and Hinton EM paper

(Optional) Murphy, Ch 11

Lecture 22:

- Semi-Supervised Learning
  - Mixture Models
  - Graph Regularization
  - Co-training

Readings: Combining Labeled and Unlabeled Data with Co-Training by Michell & Blum

Lecture 23:

- Dimensionality reduction
- Principal Component Analysis (PCA)

Readings: Shlens' PCA tutorial

(Optional) Murphy, 12.2-12.5