**Required**: Kevin Murphy, *Machine Learning, a Probabilistic Perspective*, 1st edition, MIT Press (2012) [author disavows the eBook version sold on Amazon, only buy an eBook from MIT Press]

**Supplemental:** Christopher Bishop, *Pattern Recognition and Machine Learning,* Springer (2011 printing)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Weekly Schedule** | | | | |
| **Week** | **Topics** | **Learning Goal** | **Evidence Required to Demonstrate Goals Are Achieved.**  **Be precise!!** | **Learning Activities That Will Produce the Evidence (quizzes, coding homework, etc.)**  **Be precise!! If it is a quiz, give some sample questions. If it is a coding assignment, give the specifications for the inputs, outputs, and algorithms to be implemented.** |
| 1 | Overview of Machine Learning: history, relation to classical statistics **Text:** Chapter 1; |  |  |  |
| 2 | **Linear Regression:** model specification issues, maximum likelihood estimation, ridge regression **Text:** Chapter 7, except 7.4 and 7.6 |  |  |  |
| 3 | **Logistic Regression:** model fitting methods (steepest descent, Newton’s, L1 and L2 regularization, etc.) **Online learning algorithms**: structured prediction, regret minimization, stochastic optimization, risk minimization, LMS algorithm, perceptron algorithm, relation to Bayesian techniques. **Text:** Chapter 8.1-8.3.6, 8.5, 13.3, |  |  |  |
| 4 | **Support Vector Machines:** uses in regression and classification, optimization, choice of parameter C, probabilistic interpretation **Kernel Methods:** RBF kernels, Mercer kernels, Matern kernels, Linear kernels, String kernels, Kernels derived from probabilistic generative models; the “kernel trick” and its uses: nearest neighbor classification, K-medoids clustering, ridge regression, and PCA **Text:** Chapter 14.1, 14.2, 14.5, supplemental notes by Ng |  |  |  |
| 5 | **Decision Trees:** basis in information theory, when appropriate to use, growing and pruning, detecting and avoiding over-fitting; pros and cons with respect to other ML techniques; random forests; ID3 and related algorithms; use of bias; training with incomplete data **Boosting:** why it works so well, specific types of boosting, ensemble learning **Text:** Chapters 2.8, 16.2, 16.4, 16.6 , supplemental notes by Mitchell, Yoav & Schapire |  |  |  |
| 6 | **Deep Learning:** background in feed-forward neural networks (multilayer perceptrons) and Restricted Boltzmann machines; deep generative models (sigmoid networks, Boltzmann machines, belief networks) **Text:** Chapters 16.5, 27.7, 28.1-28.2 |  |  |  |
| 7 | **Deep Learning (continued):** training of deep networks, applications of deep networks **Midterm Review** **Text:** Chapters 28.3-28.4 |  |  |  |
| 8 | **Midterm Exam Clustering:** measures of dissimilarity; k-means clustering, clustering in mixture models **Text:** Chapters 25.1, 11.1-11.3 |  |  |  |
| 9 | **Expectation Maximization:** basic idea, Jensen’s inequality, the EM algorithm, EM applied to various distributions (mixture of Gaussians), theoretical basis **Text:** Chapters 11.4, supplemental notes by Ng |  |  |  |
| 10 | **Dimensionality Reduction:** classical Principal Components Analysis, PCA theorem; Singular Value Decomposition, Probabilistic PCA, EM algorithm for PCA, PCA for categorical data, PCA for paired and multi-view data **Text:** Chapters 12.2, 12.4, 12.5 |  |  |  |
| 11 | **Graphical Techniques:** Bayesian nets, Directed Gaussian Models (DGM), plate notation, d-separation, Markov random fields, Hammersley-Clifford theorem and its effects, training maxent models, pseudo-likelihood, stochastic maximum likelihood learning **Text:** Chapter 10, 19.1-19.5 |  |  |  |
| 12 | **Graphical Techniques (continued):** belief propagation (serial, parallel, Gaussian, and other variants), variable elimination, junction trees, intractability of exact inference, approximate inference methods, max sum and max product **Text:** Chapter 20 |  |  |  |
| 13 | **Structured Prediction:** Markov and Hidden Markov (HMM) models, inference algorithms (forward, Viterbi, backward filtering, etc.), training methods, Baum-Welch algorithm, model selection, Conditional Random Fields (CRF), label-bias problem, uses in handwriting recognition, protein analysis, and stereopsis **Text:** Chapters 17.1-17.5, 19.6-19.7, CRF tutorial paper by Sutton & McCallum |  |  |  |