

Syllabus: Introduction to Machine Learning (CS189/289A)

(approximate)

- Introduction: applications, methods, concepts
- Good Machine Learning hygiene: test/training/validation, overfitting
- Linear classification
 - Perceptron algorithm
 - Support vector machines (SVMs)
- Statistical learning background
 - Decision theory; Bayes risk
 - Probabilistic models vs no model
 - Generative and discriminative models
 - Controlling complexity: regularization, bias-variance trade-off, priors.
 - Resampling, cross-validation.
 - The multivariate normal distribution.
- Linear regression
 - Least squares
 - Regularization: ridge regression, lasso
- Brief primer on optimization
- Linear Classification, revisited
 - Logistic regression
 - Linear Discriminant Analysis
 - Support vector machines revisited
 - Algorithms
 - The kernel trick
- Theoretical analysis of machine learning problems and algorithms
 - Generalization error bounds
- Nearest neighbor methods
 - k-nearest-neighbor
 - Properties of high-dimensional spaces
 - Distance learning
 - Efficient indexing and retrieval methods
- Decision trees
 - Classification and regression trees
 - Random Forests
 - Boosting
- Neural networks
 - Multilayer perceptrons
 - Variations such as convolutional nets
 - Applications
- Unsupervised methods
 - Clustering
 - Density estimation
 - Dimensionality reduction
- Applications in Data Mining

- Collaborative filtering
- The power and the peril of Big Data