Review for Exam 1

Topics

Nearest neighbor classifiers
- Nearest neighbor rule
- Role of distance function
- Training vs test error rates
- \(k\)-nearest neighbor classifiers {Homework 1 problem 1}
- Distance functions, features
- Computational issues {Homework 1 problem 1}

Predictions
- Coin toss model
- Galton board model
- Loss functions
- Optimal predictions {Practice problem 6a}
- Plug-in principle
- Maximum likelihood estimation {Homework 1 problem 2}
- Binomial tail probabilities
- Near optimality MLE + plug-in

Generative models for classification
- Optimal prediction functions (for classification)
- Generative model + Bayes' rule
- MLE + plug-in approach for generative models {Homework 1 problem 3}
- Decision boundaries
- Naive Bayes {Homework 1 problem 3}
- Generative models with Gaussian class conditionals {Practice problem 7}
- Non-parametric class conditional distributions

Risk estimation, model selection/averaging
- Risk estimation for classification
- Binomial tail bounds, likely deviation bounds
- Role of test set {Homework 2 problem 3}
- Confusion tables, ROC curves
- Model selection methods
- Pitfall of adaptivity {Homework 2 problem 3}
- Benefits of (uniform) model averaging {Practice problem 11}

Linear regression
- Optimal prediction functions (for regression) {Practice problem 6b}
- Linear regression statistical models \{Practice problem 2a\}
- Feature expansion \{Practice problem 10\}
- MLE for linear regression / ERM with squared loss \{Practice problem 2b\}
- Geometric interpretation of ordinary least squares \{Practice problem 5\}
- Plug-in interpretation of ordinary least squares \{Practice problem 8\}
- Risk and empirical risk of ERM \{Practice problem 3\}
- Fixed-design analysis \{Homework 2 problem 2\}
- Over-fitting
- Risk estimation for regression
- Inductive bias
- $\ell_2$-regularization / ridge regression
- $\ell_1$-regularization / Lasso
- Sparsity \{Homework 2 problem 1\}

**Logistic regression and linear classifiers**
- Logistic regression statistical model
- MLE for logistic regression
- Relation between logistic regression models and generative models
- Geometric of linear classifiers \{Practice problem 1a\}
- Decision boundaries (with feature expansion)
- Features for text, sparse representations
- Intractability of ERM with zero-one loss
- Linear separability
- Linear programming approach to finding linear separators
- Perceptron \{Practice problem 1b\}
- Margins and Perceptron convergence theorem
- Online Perceptron
- Online-to-batch conversion and risk bound \{Practice problem 4\}

**Support vector machines**
- Derivation of SVM (primal) problem
- Derivation of SVM dual problem
- Complementary slackness, support vectors
- Kernels \{Practice problem 13\}
- Combining kernels \{Practice problem 9\}
- Kernelization (ridge regression, Perceptron)
- Randomized Fourier-based approximation \{Practice problem 12\}
- Soft-margin SVM
- Hinge loss

**Big ideas**
- Statistical model for predictions and prediction functions
- Optimal predictions and optimal prediction functions
- The plug-in principle (e.g., model parameters, empirical distribution)
- Risk and empirical risk
- Decision boundaries
- Linear functions and feature expansions
- Inductive bias
- Aggregation (e.g., model averaging, online-to-batch)
- Mathematical optimization as a tool to formulate learning methods