STAT 32950-24620 Final, 10 am Tuesday, May 23

Final Exam Topics

The in-class final exam will be cumulative, with an emphasis on the topics covered after the midterm.

Post Midterm Topics (in addition to topics covered up to the midterm (clickable))

- Supervised learning - Classical classification

  Fisher's linear discriminants
  Linear and quadratic discriminants for normal distributions
  K-NN method

- Unsupervised learning - Cluster analysis

  K-means
  Hierarchical clustering
  Mixture Models for cluster analysis
  (likelihood function, latent membership variable approach, MLE setup, EM approach)

- Expectation - Maximization (EM) algorithm basics

  EM imputation for missing data
  EM algorithm for Maximum Likelihood Estimation
  The latent variable approach in EM

- Independent Component Analysis

  Basic assumptions and results of ICA
  Non-Gaussian-ness measures: Skewness, Kurtosis, Entropy.

- High dimensional data challenges and modern approaches

  Regularized regression Methods: Ridge, Lasso, ElasticNet regression
  (bias-variance trade-off in regularized regression models)

  Sparse PCA (using regularized regression methods)
  Support Vector Machine (SVM) for classification (introduction only, not including SVM kernel method)

- (Required for 32950 only)
  Introduction to tree based methods: regression tree, classification tree, Random Forest.