CS-498 Data Science - Midterm Review Study Guide

Midterm objectives:

- Data science concepts, process and objectives
  - Technological convergence
  - Scientific method
  - Data Science process
- Exploratory data analysis and visualization
  - Data types – numeric, categorical, ordinal
  - Data manipulation (with NumPy stack)
  - Principals of visualization
  - Applying and interpreting visualization to augment numerical analysis
- Probability and statistical inference
  - Sampling, distributions
  - Measures of central tendency
  - Variance, standard deviation
- Linear regression, multivariate analysis
  - Coefficient/factor analysis for linear regression
  - Polynomial regression
  - RSS, R^2, MSE, RMSE
- Machine learning, supervised learning
  - Machine learning concepts
  - KNN
  - Logistic regression
  - Cross-entropy loss
  - Odds, log odds, coefficient/factor analysis for logistic regression
- Classification model evaluation and metrics
  - Confusion matrix
  - Accuracy, precision, recall/sensitivity, specificity
  - TPR, FPR, ROC curve