ME 781 Statistical Machine Learning and Data Mining-Outline

1. Introduction to Machine Learning
1.1. Types of Learning Algorithms - Supervised, Unsupervised, Reinforcement Learning
1.2. Modeling Basics - Dependent and Independent Variables
1.3. Data Type and Data Scale
2. Mathematical preliminaries
2.1. Set theory
2.2. Elementary Probability and Statistics
2.3. Linear algebra for machine learning
2.4. Dissimilarity and similarity measures
3. Regression - I
3.1. Simple Linear Regression - Estimating the Coefficients
3.2. Assessing the Accuracy of the Coefficient and the estimates
3.3. Concept of t-statistic and p-value
4. Regression – II
4.1. Multiple Linear Regression
4.2. Hypothesis testing in multi-linear regression (F- statistic)
4.3. Incorporating Qualitative Variables in Modeling
4.4. Non-linear regression
4.5. Potential Problems of Linear Regression
5. Regression – III
5.1. k-Nearest Neighbors regression
5.2. Linear Regression vs KNN
5.3. Modeling optimization - Hyper-parameter tuning
6. Assessing Model Accuracy
6.1. Loss functions
6.2. Resampling Methodologies
6.2.1.Cross-Validation (The Wrong and Right Way)
6.2.2.Bootstrap
6.3. Capacity, Overfitting, and Underfitting
6.4. The Bias-Variance Trade-Off
7. Feature selection and regularization
7.1. Subset selection
7.2. Shrinkage Methods (Ridge and Lasso Regression)
7.3. High dimensional data and the curse of dimensionality
7.4. Dimensionality Reduction (PCA)
8. Non-linear models
8.1. Polynomial regression
8.2. Step functions
8.3. Splines and local regression
8.4. Generalized additive models
9. Classification – I
9.1. Simple Logistic Regression - Basic assumptions, evaluating model performance, performing predictions
9.2. Multiple Logistic Regression

9.3. Evaluating the performance of a classifier – ROC Curve, [Precision, Recall], Confusion Matrix [TP, FP, TN, FN]
9.4. Multi-class classification using Logistic Regression
9.5. Linear Discriminant Analysis
9.6. Baye's Classifier
10. Classification – II
10.1. Decision Trees – I
10.2. Classification Trees - Fitting a tree, Performing predictions
10.3. Pruning Trees
10.4. Regression Trees
10.5. Random Forests
10.6. Gradient boosting algorithms
11. Classification - III
11.1. Support Vector-Theory (Maximal Margin Classifier, SVC, SVM)
11.2. SVM using different kernels - linear, polynomial, radial
11.3. Multi-class SVM
11.4. SVM and Logistic Regression
12. Unsupervised learning
12.1. Association Rules
12.2. k-Means clustering
12.3. Hierarchical clustering
12.4. Self-Organizing Maps
12.5. PCA and Independent Component Analysis
13. Universal Approximator
13.1. Projection Pursuit Regression
13.2. Neural Networks
13.3. Forward and Backpropagation Algorithms
13.4. Introduction to Deep Learning (CNN and RNN)
14. Time series analysis
14.1. Time series forecasting
14.2. Introduction to autoregressive and moving average models
14.3. Change-point detection
15. Big Data
15.1. Introduction to big data
15.2. Big data technologies
Reference books:
1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani: An Introduction to Statistical Learning
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman: The Elements of Statistical Learning Data Mining, Inference, and Prediction
3. Ian Goodfellow, Yoshua Bengio, Aaron Courville: Deep Learning
4. Kevin P. Murphy: Machine Learning A Probabilistic Perspective
5. Thomas A. Runkler: Data Analytics_ Models and Algorithms for Intelligent Data Analysis
6. Alan F. Karr: Probabiliy, Springer-Verlag, 1993