BAN 5743 Midterm Exam Review Sheet

1. Time Series Analysis and Forecasting
2. Pattern Discovery – Non-hierarchical
   1. Advantages/ Disadvantages of Hierarchical/Non-Hierarchical
   2. K-means
   3. Transformation and Outlier Detection
   4. Profiling Hierarchical Clusters
      1. Conduction
      2. Interpretation
3. Metrics for Model Assessment
   1. Predictive model performance
   2. Business metrics
      1. Deciles
      2. Gains and cumulative gains
      3. Lift and cumulative lift
   3. Statistical/ Data science metrics
      1. ASE, misclassification, hit-ratio, sensitivity, specificity
      2. Recall, Precision, F1 score
      3. ROC curve and AUC, K-S Statistic
   4. Data Exploration and Preparation
      1. Summary statistics/ Plots
   5. Missing values
      1. Unexpected values
      2. Missing versus empty
      3. Flags
      4. Impact of missing values
      5. Choices for Imputation
   6. Transformations
      1. Types of transformations
   7. Handling extreme values
      1. Outlier detection
      2. Change or delete
4. Decision Trees
   1. Modeling Essentials
      1. Predict New Cases - rules
      2. Select useful inputs
         1. split search
         2. Logworth
      3. Optimize complexity
         1. pruning
         2. Validation assessment – accuracy, misclassification, concordance, discordance, squared error
         3. Sensitivity/ Specificity
   2. Split Criteria
      1. Purity
      2. Gini
      3. Entropy
   3. Regression Tree
      1. Interval targets
      2. Metrics
      3. F-stat and probability
   4. Pros and cons
   5. Multiple trees
      1. Gradient boosting
      2. Random forest
5. Regression Models
   1. Predict new cases - Prediction formula
      1. Manage, interpret, handle values, nonnumeric inputs, nonlinearities
   2. Linear and logistic regression
   3. Optimize complexity
      1. Best model from sequence
      2. Odds ratio
      3. Doubling Amounts
      4. Transformations and dummy coding
      5. Polynomial regression
6. Neural Networks
   1. Model essentials
   2. Multilayer Perceptron
   3. Transfer functions
   4. Learning
   5. Predict New Cases- Prediction Formula
      1. Manage, interpret, handle values, nonnumeric inputs, nonlinearities
   6. Optimize Complexity – Stopped Training
   7. Radial Basis Function Neural model
      1. Functions and shape
      2. Node placement
7. Model assessment, model implementation and special topics
   1. Model Implementation
      1. scoring
   2. Ensemble/ Surrogate Models
   3. Variable reduction vs variable selection
   4. Consolidating categorical inputs
   5. Naïve Bayes Classifier
   6. Association Rules
      1. Market Basket Analysis
      2. Association Rules
8. Text Analytics
   1. Importing text data
   2. Text parsing
   3. Term by document matrix
   4. SVD, LSA and LSI
   5. Text cluster
   6. Text topic
9. Advanced Text Analytics & Sentiment Analysis
   1. Rule Builder
   2. Predictive Modeling with text and numeric data
   3. Sentiment Analysis
10. Deep Learning Neural Network
    1. Building Blocks
    2. Training Issues
    3. Training Efficiency
    4. Deep Learning Basics
    5. Autoencoders
11. Convolutional Neural Network
    1. CNN Filter
    2. Padding and Pooling
    3. Skip Layer and Design Strategy
12. Recurrent Neural Network
    1. Applications and mechanics
    2. Vanishing gradient
    3. Gated Recurrent Unit & Long Short-term Memory
    4. Word embedding