CIS 635 Exam 1 Study Guide

1. data
   a) types - focus on nominal, ordinal and numeric
   b) quality - measurement & collection errors, noise, outliers, missing values and how to adapt
   c) preprocessing - aggregation, sampling, dim reduction (general ideas) and discretization
   d) also curse of dim.
   e) skip distance/similarity measures for now
2. exploring
   a) summary stats - mean, median, mode, range, sd
   b) viz - scatter plot, boxplot, histogram
3. classification
   a) know the general idea of using training data to create models and using test data to evaluate them
   b) confusion matrix, (page 296) note: predicted *should* be on the left and actual on the top (unlike the book)
   c) precision, recall (also page 296)
   d) curse of dimensionality – understand the general concept
   e) overfitting – this is the phenomenon of creating a model so complex that while it fits the training data well, it doesn't generalize well (doesn't fit test data well). how does it apply to techniques (think about it) (oh, yeah – it doesn't apply to naive Bayes)
4. decision trees
   a) be able to use a tree to classify instances
   b) understand the greedy algorithm generally
   c) be able to calculate gini on nodes and branches (nominal attributes only)
   d) understand how to select best split
5. naive Bayes
   a) know the Bayes formula
   b) know how to calculate priors and likelihoods from the data
   c) be able to use model to predict test data (nominal attributes only)
6. ANN
   a) know how to use a ANN with weights to predict test data
   b) understand how to use data to calculate weights (won't have to go through the entire process like in the homework)
7. SVM
   a) know how to use a model to predict test data (say you are given a line and some test samples)
   b) understand the general theory including a cursory knowledge of the kernel trick (what do different kernels allow you to do?)