CS 478 Study Guide for Midterm and Final

Bring blank paper to the tests to write your answers on. There is not room on the test for your answers. You should bring a non-programmed calculator to each test (A standard scientific calculator will suffice). Complete your answers and staple your answer sheets to the test in such a way that your writing is not covered by the staple.

The tests are closed book, but you may bring one single sided page of notes with a reasonable font (e.g. no smaller than Times 12) to the test. The spirit of this is a note page to put on equations or other items which are harder to memorize. It is not meant for trying to cram all the slides, book chapters, or knowledge from the course on a single sheet. You should know most of that without needing a sheet. Your note page will be handed in with the test. You should be prepared to answer questions from the following topic lists:

Topics from the 1st half of the semester
Perceptron
Delta Rule
Linear separability and linear models with non-linear feature preprocessing – specifically the quadric machine
Linear regression
Logistic regression
Inductive Bias, need for Bias, No free lunch
Overfit – what causes it and how to prevent it
Predicting future accuracy (N-fold CV, etc.)
MLP with Backpropagation, learning, parameter selection, etc.
Features: Approaches for selection, representation, skew, normalization and reduction
PCA
Handling missing/unknown data
Decision Trees, ID3

Final
The final is comprehensive with heavy emphasis on topics covered since the midterm.

Data Mining Process Model/Cycle (just high level)
K-Nearest Neighbor algorithm (including distance weighted, regression, reduction techniques, strengths and weaknesses)
RBF networks
Clustering approaches (K-means, HAC)
Bayesian learning (Bayes rule, MAP and ML hypotheses, Bayes optimal classifier, Naïve Bayes)
Reinforcement Learning (especially Q-learning)
Ensembles (Bagging, Boosting, Stacking, overall pros and cons)
Performance measures: Permutation testing, Precision vs Recall and ROC curves
Genetic Algorithms (Basic algorithm, data representation, genetic operators, and parameter variations)