CS498/CS698 Machine Learning - Spring 2013
Homework Assignment 3
Due at the start of class on April 11^{th}

1. (Support Vectors) For an SVM, if we remove one of the support vectors from the training set, does the size of the maximum margin decrease, stay the same, or increase for that dataset? Why? Also justify your answer by providing a simple dataset (no more than 2-D) in which you identify the support vectors, draw the location of the maximum margin hyperplane, remove one of the support vectors, and draw the location of the resulting maximum margin hyperplane.

2. (Non-linear kernels) Consider the 2-bit XOR problem for which the entire instance space is as follows:

y	x_1	x_2
-1	-1	-1
+1	-1	1
+1	1	-1
-1	1	1

   These instances are not linearly separable, but they are separable with a polynomial kernel. Recall that the polynomial kernel is of the form $K(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i x_j} + \mathbf{c})^{\mathbf{d}}$ where c and d are integers. Select values for c and d that yield a space in which the instances above are linearly separable. Write down the mapping Φ to which this kernel corresponds, write down $\Phi(x)$ for each instance above, and write down the parameters of any hyperplane in the expanded space that perfectly classifies the instances.

3. (SVM tools) In this exercise you will get some experience with a state-of-the-art SVM package, and will compare the performance of various kernels. You will need to do the following:

   - Download SVMlight from `http://www.cs.cornell.edu/People/tj/svm_light/`. This web page has links to source code and binaries, and contains information on how to build and run the various SVMlight tools.
   - Download the ionosphere dataset from the course web page. This dataset comes from the UC Irvine Machine Learning Repository. The file ionosphere.names describes the contents of the dataset. The file ionosphere.data contains the actual data in a format that is accepted by SVM-light. You'll need to manually divide the data into training and testing sets. One split will suffice.
   - Are the data linearly separable? Describe how you used SVMlight to determine this.
   - Compare the performance of the SVM using each of the following kernels - linear, polynomial, radial basis. The polynomial and radial basis kernels each take parameters. Experiment with various values of these parameters and C, which controls the tradeoff between training error and margin. Write a short report detailing your experiments. Be sure to report the margin and number of support vectors for various settings of the parameters. What kernel performed the best? What does this say, if anything, about the data?

4. (Bias-Variance) For this question you will experiment with bagging and boosting. These algorithms are available in Weka under the "Meta" category. In both cases, use the J48 decision tree algorithm as the base classifier and set numIterations to 30. Set the J48 "unpruned" option to true. For boosting, set the weightThreshold to 1000.

   To use J48, after choosing bagging (or boosting) from the "Meta" category, click on the text that starts "Bagging -P 100" beside the "Choose" button. Then click the "Choose" button beside "classifier" on the window that pops up and select J48. You should now see "J48 -C 0.25 -M 2" beside the "Choose" button. Now click that text to set the "unpruned" option.

Choose a dataset provided by Weka (one that has more than 100 instances) and compare the performance of J48, bagging, and boosting in terms of classification accuracy. What are the effects of bagging and boosting, and how can they be explained in terms of bias and variance? Now turn pruning back on in J48 and run the same experiment. Have the results changed? Explain why or why not, making a connection again to bias and variance.