Homework 4
CS 436/580L: Introduction to Machine Learning

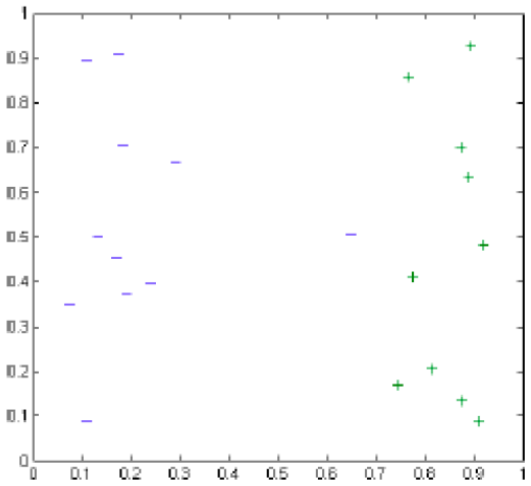
Instructor: Arti Ramesh

1 Support vector machines. (30 points)

- Download LIBSVM, currently the most widely used SVM implementation. Peruse the documentations to understand how to use it. (If you are comfortable with Weka, you can also use LIBSVM with WEKA. However, you may have to modify the data format so that WEKA is able to use it as input).
- Download the new promoters dataset in the LIBSVM format. (Available on myCourses).
- Run LIBSVM to classify promoters. Try three different kernels and use default parameters for everything else. How does it vary with different choice of kernel?

2 SVM with Slack Variables (50 points)

For this problem, assume that we are training an SVM with a) **linear** kernel and b) **quadratic** kernel (i.e., our kernel function is a polynomial kernel of degree 2). You are given the data set presented in Figure 1. The slack penalty C will determine the location of the separating hyperplane. Please answer the following questions for both **linear** kernel and **quadratic** kernel. Give a one sentence answer/justification for each and draw your solution in the appropriate part of the Figure at the end of the problem.



1. Where would the decision boundary be for very large values of C (i.e., $C \to \infty$)? Draw on the figure above. Justify your answer.
2. For $C \approx 0$, indicate in the figure below, where you would expect the decision boundary to be? Justify your answer.
3. Which of the two cases above would you expect to work better in the classification task? Why?
4. Draw a data point which will not change the decision boundary learned for very large values of C. Justify your answer.
5. Draw a data point which will significantly change the decision boundary learned for very large values of C. Justify your answer.

3 K-Nearest Neighbors (20 points)

The table below provides a training data set containing 6 observations, 3 features, and 1 class variable.

X_1	X_2	X_3	Y
0	3	0	Red
1	0	0	Red
0	1	3	Red
0	1	2	Green
-1	0	1	Green
1	0	1	Red

Suppose we wish to use this data to make a prediction for Y when using test data X1 = X2 = X3 = 0 using KNN.

1. Compute the distance between each observation and the test data.
2. What is our test classification with K-NN if we choose $K = 1$ (that is, I want to use the nearest neighbor?s label to label the test observation)? Why?
3. What is our test classification with KNN if we choose $K = 3$? Why?
4. What is the training error when $K = 1$? Suggest a method for choosing K given a training data set.