Support vector machines. (50 points)

• Download LIBSVM, currently the most widely used SVM implementation. Read 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 promoter's dataset in the LIBSVM format. (Available on the elearning).
• Run LIBSVM to classify promoters. Try different kernels (0-3) and use default parameters for everything else. How does it vary with different choice of kernel?
• Modify your perceptron classifier from homework 3 to apply to this dataset. How does the accuracy compare to what can be obtained using LIBSVM? Why? (If you are not confident about the correctness of your perceptron implementation, use the one in WEKA).

EM algorithm (50 Points)

• Download the data from the class web page.
• Implement the EM algorithm for general Gaussian mixture models (assume that the data is an array of long doubles). Use the algorithm to cluster the given data Remember that the data is 1-D. The recommendation is that you run the algorithm multiple times from a number of different initialization points (different θ^0 values) and pick the one that results in the highest log-likelihood (since EM in general only finds local maxima). One heuristic is to select r different randomly-chosen initialization conditions. For example, for each start, select the initial K Gaussian means by randomly selecting K initial data points, and select the initial K variances as all being some multiple of the overall data variance – the selection of initial covariance is not as critical as the initial means. Another option for initialization is to randomly assign class labels to the training data points and then calculate θ^0 based on this initial random assignment. Another option is to begin the iterations by executing a single M-step.

Report the parameters you get for different initializations. What initialization strategy did you use? How sensitive was the performance to the initial settings of parameters.

Now assume that variance equals 1.0 for all the three clusters and you only have to estimate the means of the three clusters using EM. Report the parameters you get for different initializations. Which approach worked better, this one or the previous one, Why? Explain your answer.

What to turn in for this part:
• Your code. EM for general GMMs and EM for GMMs with known variance.
• A report containing answers to the questions above.