1. Let's explore the maximal margin classifier on a toy data set. We are given $n = 7$ observations in $p = 2$ dimensions. For each observation, there is an associated class label Y.

Obs.	X1	X2	Y
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue

(a) Sketch the observations.
(b) Sketch the optimal separating hyperplane.
(c) Provide the equation for this hyperplane. Describe the classification rule. It should be something along the lines of ?Classify to Red if $\beta_0 + \beta_1 X1 + \beta_2 X2 > 0$.
(d) On your sketch, indicate the margin for the maximal margin hyperplane.
(e) Indicate the support vectors for the maximal margin classifier.
(f) Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane.

2. In this problem, you will use support vector approaches in order to predict Purchase based on the OJ data set.

(a) Create a training set containing a random sample of 800 observations, and a test set containing the remaining observations.
(b) Fit a linear SVM to the training data using cost=0.01, with Purchase as the response and the other variables as predictors. Describe the results obtained.
(c) What are the training and test error rates?
(d) Tune the linear SVM with various values of cost. Report the cross-validation errors associated with different values of this parameter. Select an optimal cost. Compute the training and test error rates using this new cost value. Comment on your findings.
(e) Now repeat (d), with radial basis kernels, with different values of gamma and cost. Comment on your results. Which approach seems to give the better results on this data?
(f) Now repeat again, with polynomial basis kernels, with different values of degree and cost. Comment on your results. Which approach (kernel) seems to give the best results on this data?
(g) Perform gradient boost (using gbm function in R) on the training set with 1,000 trees for a chosen values of the shrinkage parameter. You may experiment with a range of values of the shrinkage parameter.
(h) Which variables appear to be the most important predictors in the boost model?
(i) Use the boosting model to predict the response on the test data. Form a confusion matrix. How does this compare with the result SVM obtained?