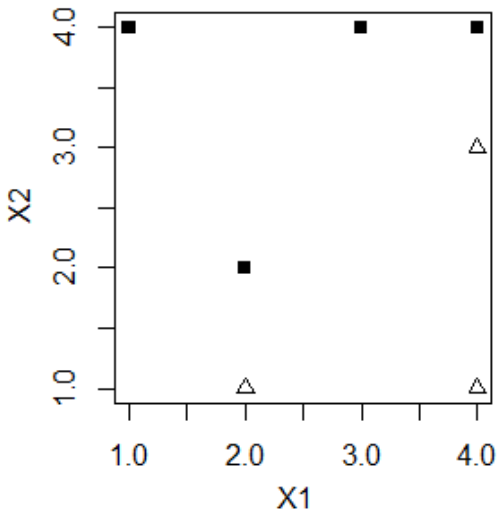
# AAE 722 Homework III

Due August 3 (Monday)

1. Here we explore the maximal margin classifier using the data set listed in the table below, which has 7 observations in 2 dimensions that belong to two classes.

| Obs. | $X_1$ | $X_2$ |
|---|---|---|
| 1 | 3 | 4 |
| 2 | 2 | 2 |
| 3 | 4 | 4 |
| 4 | 1 | 4 |
| 5 | 2 | 1 |
| 6 | 4 | 3 |
| 7 | 4 | 1 |

(1) (4 pts) Sketch the optimal separating hyperplane in the figure below and provide the equation for this hyperplane in the form of $\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$.




***Answer:*** The slope of the hyperplane is 1 and goes through the point (2,1.5). So the hyperplane is $0.5 - X_1 + X_2 = 0$.

(2) (2 pts) Describe the classification rule for the maximal margin classifier. It should be something along the lines of "Classify to ■ if $\cdots$, and classify to △ otherwise."

***Answer:*** Classify to ■ if $0.5 - X_1 + X_2 > 0$, and classify to △ otherwise

(3) (2 pts) On the figure, indicate the margin for the maximal margin hyperplance. Calculate the margin. Note that the distance of the point $(x_0, y_0)$ to the line $ax + by + c = 0$ is $\frac{|ax_0+by_0+c|}{\sqrt{a^2+b^2}}$.

***Answer:*** The margin is: $1/(2\sqrt{2})$.

(4) (2 pts) What are the support vectors for the maximal margin classifier?

***Answer:*** The support vectors are (2,1), (2,2), (4,3), and (4,4).

2. (10 pts) Use the `OJ` data set, which is part of the `ISLR` package, to answer the following questions.

(1) Create a training set containing a random sample of 900 observations, and a test set containing the remaining observations.
(2) Fit a tree to the training data, with `Purchase` as the response and the other variables as predictors. Use the `summary` function to produce summary statistics about the tree, and describe the results obtained. What is the training error rate? How many terminal nodes does the tree have?
(3) Type in the name of the tree object in order to get a detailed text output. Pick one of the terminal nodes, and interpret the information displayed.
(4) Create a plot of the tree, and interpret the results.
(5) Predict the response on the test data, and produce a confusion matrix comparing the test labels to the predicted test labels. What is the test error rate?
(6) Apply the `cv.tree()` function to the training set in order to determine the optimal tree size.
(7) Produce a plot with tree size on the $x$-axis and cross-validated classification error rate on the $y$-axis.
(8) Which tree size corresponds to the lowest cross-validated classification error rate?
(9) Produce a pruned tree corresponding to the optimal tree size obtained using cross-validation.
(10) Compare the training error rates between the pruned and unpruned trees. Which is higher?
(11) Compare the test error rates between the pruned and unpruned trees. Which is higher?

3. (10 pts) Based on the `Auto` data set, predict whether a given car gets high or low gas mileage.

(1) Create a binary variable that takes on a 1 for cars with gas mileage above the median, and a 0 for cars with gas mileage below the median.

(2) Fit a support vector classifier to the data with various values of `cost`, in order to predict whether a car gets high or low gas mileage. Report the cross-validation errors associated with different values of this parameter. Comment on your results.

(3) Now repeat (2), this time using SVMs with radial and polynomial basis kernels, with different values of `gamma` and `degree` and `cost`. Comment on your results.

4. (10 pts) Use the `Boston` data set in the `MASS` package to predict the median value of owner-occupied homes (in \$1000's) by training the neural networks. Employ all available predictors in the data, including: (i) `crim` - per capita crime rate by town, (ii) `zn` - proportion of residential land zoned for lots over 25,000 sq.ft., (iii) `indus` - proportion of non-retail business acres per town, (iv) `chas` - Charles River dummy variable (1 if tract bounds river; 0 otherwise), (v) `nox` - nitric oxides concentration (parts per 10 million), (vi) `RM` - average number of rooms per dwelling, (vii) `age` - proportion of owner-occupied units built prior to 1940, (viii) `dis` - weighted distances to five Boston employment centers, (ix) `rad` - index of accessibility to radial highways, (x) `tax` - full-value property-tax rate per \$10,000, (xi) `ptratio` - pupil-teacher ratio by town, (xii) `black` - $1000(Bk - 0.63)^2$ where $Bk$ is the proportion of blacks by town, (xiii) `lstat` - % lower status of the population.

(1) When training your network, vary the number of hidden layer from 1 to 10 but keep the number of neurons in each layer at 5. Find the best network that has the lowest MSE. Remember to apply min-max scale to the data before fitting the network.Also split the data to 75% training and 25% test data.
(2) Do 10-fold cross validation on the network determined in (1) and report the mean CV error.