Homework 2

Stat 435, Spring 2020

Due Friday, April 24, 11:59pm

Linear regression

If you are not familiar with R you may want to go through Chapter 3.6: "Lab: Linear Regression" of ISLR.

Problem 1 (25 points)

In this problem, we will make use of the `Auto` data set, which is part of the `ISLR` package.

1. Fit a least squares linear model to the data, in order to predict `mpg` using all of the other predictors except for `name`. Present your results in the form of a table. Be sure to indicate clearly how any qualitative variables should be interpreted. For each predictor, comment on whether you can reject the null hypothesis that there is no linear association between that predictor and gas mileage, conditional on the other predictors in the model.
2. What is the resubstitution (training) mean squared error of this model?
3. What gas mileage do you predict for a Japanese car with three cylinders, displacement 100, horsepower of 85, weight of 3000, acceleration of 20, built in the year 1980?
4. On average, holding all other covariates fixed, what is the difference between the `mpg` of a Japanese car and the `mpg` of an American car?
5. On average, holding all other covariates fixed, what is the change in `mpg` associated with a 10-unit change in horsepower?

Problem 2 (25 points)

Consider using only the `origin` variable to predict `mpg` on the `Auto` data set. In this problem, we will explore the coding of this qualitative variable.

1. First, code the `origin` variable using two dummy (indicator) variables, with `Japanese` as the default value. Write out an equation like (3.30) in the textbook, and report the coefficient estimates. What is the predicted `mpg` for a Japanese car? for an American car? for a European car?
2. Now, code the `origin` variable using two dummy (indicator) variables, with `American` as the default. Write out an equation like (3.30) in the textbook, and report the coefficient estimates. What is the predicted `mpg` for a Japanese car? for an American car? for a European car?

3. Now, code the `origin` variable using two variables that take on values of $+1$ or -1, as in the top of page 85 of the textbook. Write out an equation like (3.30) in the textbook, and report the coefficient estimates. What is the predicted `mpg` for a Japanese car? for an American car? for a European car?

4. Finally, code the `origin` variable using a single variable that takes on values of 0 for Japanese, 1 for American, and 2 for European. Write out an equation like (3.30) in the textbook, and report the coefficient estimates. What is the predicted `mpg` for a Japanese car? for an American car? for a European car?

5. Comment on your results in (a)-(d).

Problem 3 (20 points)

Consider using least squares linear regression to predict weight (Y) using height.

1. Suppose that you measure height in inches (X_1), fit the model

$$Y = \beta_0 + \beta_1 X_1 + \epsilon,$$

and obtain the coefficient estimates $\hat{\beta}_0 = -165.1$ and $\hat{\beta}_1 = 4.8$. What weight will you predict for an individual who is 64 inches tall?

2. Now suppose that you want to measure height in feet (X_2) instead of inches. (There are 12 inches to a foot.) You fit the model

$$Y = \beta_0^* + \beta_1^* X_2 + \epsilon.$$

What are the coefficient estimates? What weight will you predict for an individual who is 64 inches tall (i.e. 5.333 feet)?

3. Now suppose you fit the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon,$$

which contains both height in inches and height in feet as predictors. Provide a general expression for the least squares coefficient estimates for this model.

4. How do the (training set) mean squared errors compare for three models fit in (a)–(c)?