Problem Set 2 – MATH 243

Theodore Dounias

September 7, 2017

Exercises 1, 4, 5, 6; Chapter 2, ISLR

1

The p-value describes the null hypothesis that the coefficient is equal to zero, implying no relation between the predictor and the response variable.

In terms of table 3.4, both Radio and TV adds seem to have some sort of relationship with the number of sales, which based on their corresponding coefficients is positive. However the p-value for newspaper adds implies that we fail to reject the null hypothesis of no relation with sales, making it possible that newspaper adds do not impact how much people buy a product.

4

a) In this case we expect the training RSS to be lower for the cubic regression. This is true because it better matches the data regardless of the data's underlying linearity; the model would just predict relatively small coefficients for the square and cubic parameters, with higher p-values to compensate for the linearity, but would still be more accurate for the training data. The only possible problem would be that...

b) ...this model would overfit for the training data set, creating a set of larger residuals when compared to test data, which means that the linear model will probably produce a smaller RSS in this case.

c) Based on the same explanation we gave previously for (a), the RSS would be greater for the linear model.

d) In this case I would say we need to know how far from linear, and how close to cubic the relationship of X and Y is. If, for example, the data is almost linear, then the overfitting might be worse than any possible bias we get from the linear model. However, if the relation is described by a quartic polynomial, then the linear model's bias might be so extreme as to render the rate at which overfitting occures negligible.

5

Let's assume that:

$$\sum_{k=1}^{n} x_k^2 = c$$

Then, since $i \neq k$ we can write:

$$\hat{b} = (\sum_{i=1}^{n} x_i y_i)/c \Leftrightarrow \hat{b} = \sum_{i=1}^{n} (x_i/c) y_i$$

Using our initial model, we can easily conclude that:

$$\hat{y}_i = x_i \sum_{j=1}^{n} (x_j/c) y_j \Leftrightarrow \hat{y}_i = \sum_{j=1}^{n} a_j y_j$$

With $a_j = (x_i x_j/c)$. a_j here describes the weight of each value of y_j in their linear combination to output the fitted values as the corresponding x_i times each individual value of the explanatory over the squared sum of all values of the explanatory variable.

6

If we are using the model described in (3.4) and minimizing the RSS, then by definition of $\hat{b}_0$ the linear regression line will ass through $(\bar{x}, \bar{y})$.

$$\hat{b}_0 = \bar{y} - \hat{b}_1\bar{x} \Leftrightarrow \bar{y} = \hat{b}_1\bar{x} + \hat{b}_0$$