Exercises Linear Regression

Exercise 1

Question 1

- Describe the null hypotheses to which the p-values given in Table . Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of sales, TV,radio , and newspaper , rather than in terms of the coefficients of the linear model

Answer :

The **Null hypothesis** for each variable is

$$H_0 : \text{TV}=\text{Radio}=\text{Newspaper}=0$$

There is no association between the mediums and **Sales**

- The TV and $Radio$ $p-values$ are both extremely low, means a **strong evidence to reject the null hypothesis** which indicate TV and $Radio$ advertising are **Significantly associated with sales**
- For $Newspaper$ the $p-value$ is very high , means we **fail to reject the null hypothesis** which indicate that the $Newspaper$ doesn't affect **sales**

Question 2

- Carefully explain the differences between the KNN classifier and KNN regression methods

Answer

- **KNN** classifier is a method to predict the class of a **qualitative** response, **approximates** the Bayes classifier, it finds K nearest Training Data points and assigns the most common class label among them(majority vote)
- **KNN** regression its a method to predict **quantitative** responses its takes the average of the responses of the K nearest neighbors to predict the value of the observation

Question 3

- Suppose we have a data set with five predictors,$X_1 = \text{GPA}$, $X_2 = \text{IQ}$
  , $X_3 =$ **Level** (1 for College and 0 for High School), X_4= Interaction between **GPA** and **IQ**, and X_5 = Interaction between **GPA** and **Level**. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and get

$$\hat{\beta}_0 = 50, \hat{\beta}_1 = 20, \hat{\beta}_2 = 0.07, \hat{\beta}_3 = 35, \hat{\beta}_4 = 0.01, \hat{\beta}_5 = -10$$

Answer

- The **College graduates always earn more** than high school grads,provided that the **GPA** is high enough

$$\text{Salary}_{college} = 50 + 20 \times \text{GPA} + 0.07 \times \text{IQ} + 35 + 0.01 \times (\text{GPA} \times \text{IQ}) - 10 \times (\text{GPA})$$

$$\text{Salary}_{college} = 50 + 10 \times \text{GPA} + 0.07 \times \text{IQ} + 35 + 0.01 \times (\text{GPA} \times \text{IQ})$$

$$\text{Salary}_{highschool} = 50 + 20 \times \text{GPA} + 20 \times \text{IQ} + 0.01 \times (\text{GPA} \times \text{IQ})$$

$$\text{Salary}_{college-highschool} = 35 - 10 \times \text{GPA}$$

- If **GPA** is high enough **College grads** have higher salary
- Predicted **Salary** for a college grad with **IQ** of 110 and **GPA** of 4.0

$$\text{Salary} = 50 + 10.4 + 0.07.110 + 35 + 0.01 \times (4.110)$$

$$\text{Salary} = 137.1$$

- Since the **coefficient** for **GPA/IQ** interaction term us very small, there is very little evidence of an interaction effect