Contents

Linear Regression

Notes

Exercises

Conceptual

Question 1

Describe the null hypothesis to which the p-values given in Table 3.4 correspond...

The p-value for each of the predictors (intercept, TV, radio and newspaper) correspond to the null hypothesis that the coefficient estimate is NOT significantly different from 0. The small p-values suggest that we can reject the null hypothesis and that the intercept, TV and radio are all significant predictors in the model to determine product sales.

Based on the reported values we can conclude that TV and radio are important factors in determining product sales, while newspaper spending is not. Also that additional radio spending seems to contribute a greater amount to product sales than additional spending on TV.

More concretely, we conclude that the impact of an increase of \$1,000 in spending on TV, while holding other factors constant will, on average, increase product sales in units by 46.

Question 2

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

- The KNN regression compared to KNN classifier yields a numeric value rather than a categorical value in its prediction (ie quantitative vs qualitative)

Question 3

Dataset with five predictors...

Functional form

$$Y = \beta_0 + \beta_1 * GPA + \beta_2 * IQ + \beta_3 * Gender + \beta_4 * Gender * IQ + \beta_5 * GPA * Gender + \epsilon$$

Coefficients

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

a) *Which answer is correct, and why?*

   iii. For a fixed value of IQ and GPA, males earn more on average than females provided that the GPA is high enough.

The interaction term on Gender and GPA interaction (X_5) is negative, meaning that females with higher GPAs earn relatively less than their male counterparts–however that is only after accounting for the positive coefficient on female earnings (X_3)