Introduction to Statistical Learning
Chapter 3 Exercise

Penghao Chen

February 18, 2019

1. The null hypothesis is that the intercept, TV, radio and newspaper term has no effect on sales respectively, i.e., the coefficients are 0, respectively.

   From the table, because of the small p-value, it can be inferred that the intercept, TV and ratio have statistically significant effect on product sales. But for newspaper, because of its large p-value, its effect is not statistically significant.

2. The KNN classifier applies to qualitative responses. It defines a neighbourhood, then calculate the probability of each class in that neighbourhood, and classifies a point to the most-likely class.

   The KNN regression method applied to quantitative responses. It defines a neighbourhood, then calculate the average of the responses in that neighbourhood as the response to the point in question.

3. (a)

   i. False. It depends on the GPA level. When GPA is low, females earn more than males do. When GPA is high, the situation reverses.

   ii. False. It depends on the GPA level.

   iii. True.

   iv. False.

   (b) 137.1k.

   $$50 + 20 \times 4 + 0.07 \times 110 + 35 \times 1 + 0.01 \times 4 \times 110 - 10 \times 4 \times 1 = 137.1$$

   (c) False. Since we don't know the t-statistic or p-value for this interaction term, we cannot say about whether the evidence is strong or not. The magnitude of the coefficient does not imply whether the evidence is week or strong. Plus, the response is in thousand of dollars. The effect is not small.

4. (a) The training RSS for the cubic regression is lower, since the cubic regression overfits the data.

(b) The test RSS for the linear regression is expected to be lower, since the linear regression estimator is closer to the true model and has lower variance.

(c) The training RSS for the cubic regression is lower, since in the non-linearity case, the model has more flexibility.

(d) There is not enough information to tell which has lower test RSS. If the true model is closer to linear, e.g., $X^{0.5}$, then linear regression is expected to have lower test RSS. If the true model is closer to the cubic regression, e.g., X^4, then cubic regression is expected to have lower test RSS.

5. Substitute the expression of β into $\hat{y_i} = \hat{x_i}\beta$, we get:

$$\begin{aligned}\hat{y_i} &= x_i \frac{\sum_{i'=1}^{n} x_{i'} y_{i'}}{\sum_{i'=1}^{n} x_{i'}^2} \\ &= \sum_{i'=1}^{n} \frac{x_{i'} x_i}{SSX} y_{i'}\end{aligned}$$

where $SSX = \sum_{i=1}^{n} x_i^2$ is a constant given x.

Hence, $a_{i'} = \frac{x_{i'} x_i}{SSX}$

6. *Proof.* Plugging in the expression for $\hat{\beta}_0$ and $(\bar{x}, \bar{y})$ into $y = \hat{\beta}_0 + \hat{\beta}_1 x$:

$$\begin{aligned}RHS &= \bar{y} - \hat{beta_0}\bar{x} + \hat{beta_0}\bar{x} \\ &= \bar{y} \\ &= LHS\end{aligned}$$

$\square$

7. *Proof.*

$$\begin{aligned}
R^2 &= 1 - \frac{RSS}{TSS} \\
&= 1 - \frac{\sum\limits_i (y_i - \hat{y}_i)^2}{\sum\limits_i y_i^2} \\
&= 1 - \frac{\sum\limits_i (y_i - \hat{\beta} x_i)^2}{\sum\limits_i y_i^2} \\
&= 1 - \frac{\sum\limits_i (y_i - \frac{\sum\limits_j x_j y_j}{\sum\limits_j x_j^2} x_i)^2}{\sum\limits_i y_i^2} \\
&= 1 - \frac{\sum\limits_i (y_i^2 + \left(\frac{\sum\limits_j x_j y_j}{\sum\limits_j x_j^2}\right)^2 x_i^2 - 2 y_i \frac{\sum\limits_j x_j y_j}{\sum\limits_j x_j^2} x_i)}{\sum\limits_i y_i^2} \\
&= 1 - \frac{\sum\limits_i y_i^2 + \frac{(\sum\limits_i x_i y_i)^2}{\sum\limits_i x_i^2} - 2 \frac{(\sum\limits_i x_i y_i)^2}{\sum\limits_i x_i^2}}{\sum\limits_i y_i^2} \\
&= \frac{(\sum\limits_i x_i y_i)^2}{(\sum\limits_i x_i^2)(\sum\limits_i y_i^2)} \\
&= \frac{(Cov(X,Y))^2}{Var(X)^2 Var(Y)^2} \\
&= \rho^2
\end{aligned}$$

□

11 (d)

Proof.

$$\begin{aligned}
t &= \frac{\hat{\beta}}{S.E.(\hat{\beta})} \\
&= \frac{\frac{\sum\limits_i x_i y_i}{\sum\limits_i x_i^2}}{\sqrt{\frac{\sum\limits_i (y_i - \hat{\beta} x_i)^2}{(n-1)\sum\limits_i x_i^2}}} \\
&= \frac{\sqrt{(n-1)}\sum\limits_i x_i y_i}{\sqrt{(\sum\limits_i x_i^2)[\sum\limits_i (y_i - \hat{\beta} x_i)^2]}} \\
&= \frac{\sqrt{(n-1)}\sum\limits_i x_i y_i}{\sqrt{(\sum\limits_i x_i^2)[\sum\limits_i (y_i^2 - 2\hat{\beta} x_i y_i + \hat{\beta}^2 x_i^2)]}} \\
&= \frac{\sqrt{(n-1)}\sum\limits_i x_i y_i}{\sqrt{(\sum\limits_i x_i^2)[\sum\limits_i y_i^2 - 2(\frac{\sum\limits_i x_i y_i}{\sum\limits_i x_i^2})(\sum\limits_i x_i y_i) + \frac{(\sum\limits_i x_i y_i)^2}{(\sum\limits_i x_i^2)^2}(\sum\limits_i x_i^2)]}} \\
&= \frac{\sqrt{(n-1)}\sum\limits_i x_i y_i}{\sqrt{(\sum\limits_i x_i^2)(\sum\limits_i y_i^2) - (\sum\limits_i x_i y_i)^2}}
\end{aligned}$$

□

12 (a) When the variance of x is equal to the variance of y. This is because: $\hat{\beta}_{y\sim x} = \frac{Cov(x,y)}{Var(x)}$ and $\hat{\beta}_{x\sim y} = \frac{Cov(x,y)}{Var(y)}$. Hence $\hat{\beta}_{y\sim x}$ equals to $\hat{\beta}_{x\sim y}$ iff $Var(x) = Var(y)$.