Stat Learning HW 2

Jay Lee

9/7/2017

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

Table 3.4

	Coefficient	Std. error	t-statistic	p-value
Intercept	2.939	0.3119	9.42	<0.0001
TV	0.046	0.0014	32.81	<0.0001
radio	0.189	0.0086	21.89	<0.0001
newspaper	-0.001	0.0059	-0.18	0.8599

- These p-values refer to the null hypotheses that (respectively) with no advertising we would see no `sales`, `TV` advertising has no effect on `sales`, `radio` advertising has no effect on `sales`, and `newspaper` advertising has no effect on `sales`.
- Because the first three estimates have a p-value < 0.05, we can conclude that without advertising we would not see zero `sales`, `TV` advertising has some effect on `sales`, and `radio` advertising has some effect on `sales`.

4. I collect a set of data($n = 100$ observations) containing a single predictor and a quantitative response. I then fit a linear regression model to the data, as well as a separate cubic regression, i.e. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$.

(a) Suppose that the true relationship between X and Y is linear, i.e. $Y = \beta_0 + \beta_1 X + \epsilon$. Consider the training residual sum of squares (RSS) for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.
(b) Answer (a) using test rather than training RSS.
(c) Suppose that the true relationship between X and Y is not linear, but we don't know how far is it from linear. Consider the training RSS for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.
(d) Answer (c) using test rather than training RSS.

- We would expect the cubic training RSS to be lower, because it would pick up more variability in the data.
- We would expect the linear test RSS to be lower, because the cubic model would pick up variability from the training data, and this variability is not necessarily present in the test data.
- We would expect the cubic training RSS to be lower, because it would pick up more variability in the data. This is true even if we don't know the pattern of the data.
- We don't know, but I would guess the cubis test RSS would be lower, because the linear model would miss any systematic curve in the data.

5. Consider the fitted values that result from performing linear regression without an intercept. In this setting, the ith fitted value takes the form

$$\hat{y}_i = x_i \hat{\beta},$$

where

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i y_i\right) / \left(\sum_{i'=1}^{n} x_{i'}^2\right).$$

Show that we can write

$$\hat{y}_i = \sum_{i'=1}^{n} a_{i'} y_{i'}.$$

What is $a_{i'}$?

Proof.

$$\begin{aligned}
\hat{y}_i &= x_i \hat{\beta} \\
&= x_i \left(\sum_{i=1}^{n} x_i y_i\right) \div \left(\sum_{i'=1}^{n} x_{i'}^2\right) \\
&= x_i \frac{x_1 y_1 + x_2 y_2 + \cdots x_n y_n}{x_1^2 + x_2^2 + \cdots x_n^2} \\
&= \frac{x_i x_1}{\sum_{i'=1}^{n} x_{i'}^2} y_1 + \frac{x_i x_2}{\sum_{i'=1}^{n} x_{i'}^2} y_2 + \cdots \frac{x_i x_n}{\sum_{i'=1}^{n} x_{i'}^2} y_n. \\
\text{Let } a_i' &= \frac{x_i x_i'}{\sum_{i'=1}^{n} x_{i'}^2}, \text{ then} \\
\hat{y}_i &= \sum_{i'=1}^{n} a_{i'} y_{i'}.
\end{aligned}$$

□

6. Using (3.4), argue that in the case of simple linear regression, the least squares line always passes through the point $(\bar{x}, \bar{y})$

Proof. We want to show that the line $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ goes through the point $(\bar{x}, \bar{y})$. Plugging in $x_i = \bar{x}$ on the right side, we get $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$. By (3.4), $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$. Then, plugging in $\hat{\beta}_0$,

$$\begin{aligned}
\hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\
&= \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} \\
&= \bar{y}.
\end{aligned}$$

Thus, $(\bar{x}, \bar{y})$ lies on the least squares line. □