1.1 [1 point] ISLR 2e (Gareth James, et al.)

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

Answer:

Let training samples be (x1,y1),...,(xn,yn), then define the residual sum of squares (RSS) as

$$\mathrm{RSS} = \left(y_1 - \widehat{\beta_0} - \widehat{\beta_1}\, x_1\right)^2 + \left(y_2 - \widehat{\beta_0} - \widehat{\beta_1}\, x_2\right)^2 + \cdots + \left(y_n - \widehat{\beta_0} - \widehat{\beta_1}\, x_n\right)^2$$

And we get

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$$

So, the least squares line always passes through the point ($\bar{x}$, $\bar{y}$)

1.2 [1 point] In HW-2, Problem 2.1(c) and (g), you created two linear regression models. For each of these models, using residual analysis (see lecture notes) analyze the predictive properties of the models. (That is, look at the values of RSS and the residual plot and comment on how well do you think your model will predict.)

Answer:

(c) The model RSS value is relatively small, and the residual distribution is relatively concentrated, indicating that the model fits better. (g) The model RSS value is relatively large, and the residual distribution is relatively scattered, indicating that the model fits relatively generally.