Homework2

Name: Chenqing Ji
Student ID:11911303
Statistical Learning for Data Science

Due time: March 20, 2023 (Monday) 12:00am

1 Proof

Proof the bias-variance trade-off in detail:
(hint: bias/variance decomposition of MSE)

Proof:

For Bias/variance decomposition of MSE, we can have:

$$\begin{aligned} \boldsymbol{E}\left(\boldsymbol{y_0}-\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)^{\mathbf{2}} &= \boldsymbol{E}\left(\boldsymbol{f}\left(\boldsymbol{x_0}\right)+\boldsymbol{\varepsilon}-\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)^{\mathbf{2}} \\ &= \boldsymbol{E}\left(\boldsymbol{f}\left(\boldsymbol{x_0}\right)-\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)+\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)-\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)+\boldsymbol{\varepsilon}\right)^{\mathbf{2}} \\ &= \boldsymbol{E}\left(\begin{array}{l} \left(\boldsymbol{f}\left(\boldsymbol{x_0}\right)-\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)\right)^{\mathbf{2}}+\left(\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)-\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)^{\mathbf{2}} \\ +\mathbf{2}\left(\boldsymbol{f}\left(\boldsymbol{x_0}\right)-\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)\right)\left(\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)-\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right) \\ +\boldsymbol{\varepsilon}^{\mathbf{2}}+\mathbf{2}\boldsymbol{\varepsilon}\left(\boldsymbol{f}\left(\boldsymbol{x_0}\right)-\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)\right) \\ +\mathbf{2}\boldsymbol{\varepsilon}\left(\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)-\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right) \end{array}\right) \end{aligned} \tag{1}$$

So, in the equation (1), there exists **6** terms. As we know, $\boldsymbol{f}\left(\boldsymbol{x_0}\right)$ can be regarded as a constant in this expression, $\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)$ and $\boldsymbol{\varepsilon}$ can be regarded as variables in this expression. It's worth noting that although $\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)$ can be regarded as variable, $\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)$ should be regarded as a constant.

Therefore, some terms can be simplified in the equation (1) like that:

$$\boldsymbol{E}\left(\ \mathbf{2}\left(\boldsymbol{f}\left(\boldsymbol{x_0}\right)-\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)\right)\left(\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)-\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)\ \right)=\mathbf{0} \tag{2}$$

$$\boldsymbol{E}\left(\ \mathbf{2}\boldsymbol{\varepsilon}\left(\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)-\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)\ \right)=\mathbf{0} \tag{3}$$

$$\boldsymbol{E}\left(\ \mathbf{2}\boldsymbol{\varepsilon}\left(\boldsymbol{f}\left(\boldsymbol{x_0}\right)-\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)\right)\ \right)=\mathbf{0} \tag{4}$$

Note that the equations (3) and (4) are true because $\boldsymbol{E}\left(\boldsymbol{\varepsilon}\right)$ is 0. Then, the equation (1) can be simplified as:

$$\boldsymbol{E}\left(\boldsymbol{y_0}-\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)^{\mathbf{2}}=\left(\boldsymbol{f}\left(\boldsymbol{x_0}\right)-\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)\right)^{\mathbf{2}}+\boldsymbol{E}\left(\left(\boldsymbol{E}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)-\hat{\boldsymbol{f}}\left(\boldsymbol{x_0}\right)\right)^{\mathbf{2}}\right)+\boldsymbol{E}\left(\boldsymbol{\varepsilon}^{\mathbf{2}}\right) \tag{5}$$

From the equation (5), the first term represent the ***Square*** of the ***Bias*** for $\hat{\boldsymbol{f}}(\boldsymbol{x_0})$. The second term represent the variance of $\hat{\boldsymbol{f}}(\boldsymbol{x_0})$ and the third term represent the variance of $\boldsymbol{\varepsilon}$, which is also called ***irreducible error***.

Therefore, due to the Bias/variance decomposition of MSE, the expected test MSE can be expressed in the equation (6) below, which is called "The Bias-Variance Trade-off".

$$\mathbf{MSE} = \mathbf{Var}\left(\hat{\boldsymbol{f}}(\boldsymbol{x_0})\right) + \left[\mathbf{Bias}\left(\hat{\boldsymbol{f}}(\boldsymbol{x_0})\right)\right]^{\mathbf{2}} + \mathbf{Var}(\boldsymbol{\varepsilon}) \tag{6}$$

2 Proof

For simple linear regression, proof:

$$\mathrm{Var}(\hat{\beta}_0) = \mathrm{SE}(\hat{\beta}_0)^2 = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

$$\mathrm{Var}(\hat{\beta}_1) = \mathrm{SE}(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Proof:

For simple linear regression: $\boldsymbol{y_i} = \boldsymbol{\beta_1 x_i} + \boldsymbol{\beta_0} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}$ is a random variable, so $\boldsymbol{y_i}$ is also a random variable.

$$\mathbf{Var}\left(\hat{\beta}_1\right) = \mathbf{Var}\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) = \mathbf{Var}\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2} - \bar{y}\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

$$\mathbf{Because}\sum_{i=1}^{n}(x_i - \bar{x}) = 0, \bar{y}\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0$$

$$\mathbf{Var}\left(\hat{\beta}_1\right) = \mathbf{Var}\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \mathbf{Var}(y_i)}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}, \mathbf{Var}(y_i) = \mathbf{Var}(\varepsilon) = \sigma^2$$

$$\mathbf{So\ Var}\left(\hat{\beta}_1\right) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \Rightarrow \mathbf{Var}\left(\hat{\beta}_0\right) = \mathbf{Var}\left(\bar{y} - \hat{\beta}_1\bar{x}\right) = \mathbf{Var}(\bar{y}) + \bar{x}^2\,\mathbf{Var}\left(\hat{\beta}_1\right) - 2\bar{x}\,\mathbf{Cov}\left(\bar{y}, \hat{\beta}_1\right)$$

$$\mathbf{Suppose}\ c_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\mathbf{Cov}\left(\bar{y}, \hat{\beta}_1\right) = \mathbf{Cov}\left(\frac{1}{n}\sum y_i, \sum c_j y_j\right) = \frac{1}{n}\sum c_j \sum \mathbf{Cov}\left(y_i, y_j\right) = \frac{1}{n}\sum c_j \,\mathbf{Cov}\left(y_j, y_j\right) = \frac{1}{n}\sum c_i = 0$$

$$\mathbf{So, Var}\left(\hat{\beta}_0\right) = \mathbf{Var}(\bar{y}) + \bar{x}^2\,\mathbf{Var}\left(\hat{\beta}_1\right) = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right]$$

3 Problem

3.7.1 Solution:

The null hypothesises are that: For "TV" , in the presence of radio ads and newspaper ads, TV ads have no effect on sales. Similarly, for "radio" , in the presence of TV and newspaper ads, radio ads have no effect on sales; For "newspaper" , in the presence of TV and radio ads, newspaper ads have no effect on sales. Then, from the p-value we can get that: The low p-values of TV and radio suggest that the null hypotheses for TV and radio are false. However, the high p-value of newspaper suggests that the null hypothesis for newspaper is true.

3.7.2 Solution:

As we all know , KNN is a non-parametric algorithm used for both classification and regression tasks. The main difference between KNN classifier and KNN regression methods is the type of output that they produce: KNN classifier method produces a categorical output, while KNN regression method produces a continuous output.

3.7.3 Solution:

(a) The relationship between the starting salary after graduation (in thousands of dollars) and related five predictors is shown in (7) and (8):

$$\boldsymbol{Salary = 50 + 20 * GPA + 0.07 * IQ + 35 * Gender + 0.01 * (GPA * IQ) - 10 * (GPA * Gender)} \tag{7}$$

It will convert to:

$$\boldsymbol{Y = 50 + 20 * X_1 + 0.07 * X_2 + 35 * X_3 + 0.01 * X_4 - 10 * X_5} \tag{8}$$

Where $\boldsymbol{X_4 = X_1 * X_2}$ and $\boldsymbol{X_5 = X_1 * X_3}$ For male, the gender is 0 and for female the gender is 1. Therefore,the equation (8) for male and female will convert to the equation (9) and (10):

$$\boldsymbol{Y = 50 + 20 * X_1 + 0.07 * X_2 + 0.01 * X_4} \tag{9}$$

$$\boldsymbol{Y = 50 + 20 * X_1 + 0.07 * X_2 + 35 + 0.01 * X_4 - 10 * X_1} \tag{10}$$

Based on the equation (7)(8)(9)(10), we can get that: Once the GPA($\boldsymbol{X_1}$)is high enough, for a fixed value of IQ and GPA, males earn more on average than females. Therefore, the answer iii. is true.

(b) If a female with IQ of 110 and a GPA of 4.0, her profile will be: $\boldsymbol{X_1 = 4}$, $\boldsymbol{X_2 = 110}$, $\boldsymbol{X_4 = X_1 * X_2 = 4 * 110}$.

Her salary will be:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{50 + 20 * 4 + 0.07 * 110 + 35 + 0.01(4 * 110) - 10 * 4} \\ &= \boldsymbol{137.1} \end{aligned} \tag{11}$$

Therefore, her salary will be **137.1** (in thousands of dollars).

(c) **False**. Because before we judge it , we should carefully check the p-value of the coefficient for the GPA/IQ interaction term to determine whether the interaction term is statistically significant or not.

3.7.4 Solution:

(a) Yes, we would expect that one to be lower than the other. That is, the cubic regression will have a lower training RSS than the linear regression because the cubic regression can make a closer fit with the data that has with a larger irreducible error.

(b) Different from (a), we expect the cubic regression to have a higher test RSS than the linear regression because the training process will overfit the cubic regression so it will have more error (RSS) than the linear regression in the test process.

(c) My answer to this problem is the same as the problem (a). That is: the cubic regression will have a lower training RSS than the linear regression. Because the cubic regression still has higher flexibility than linear regression no matter what the true relationship between X and Y is. Then, it will fit the training points closer than the linear regression and reduce the training error (RSS).

(d) It is a shame that there is not enough evidence to tell which test RSS will be lower between the linear regression and the cubic regression because we don't know how far it is from linear for the true relationship between X and Y. If the true relationship between X and Y is closer to linear than cubic, the test RSS for linear regression could be lower than the related cubic regression. However, if the true relationship between X and Y is closer to the cubic than linear, the test RSS for the cubic regression could be lower than the related linear regression.

3.7.5 Solution:
For this problem, we have the known information about equation (12) and (13):

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \boldsymbol{x}_{\boldsymbol{i}}\hat{\boldsymbol{\beta}} \tag{12}$$

$$\hat{\boldsymbol{\beta}} = \left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x}_{\boldsymbol{i}}\boldsymbol{y}_{\boldsymbol{i}}\right) / \left(\sum_{\boldsymbol{i'=1}}^{\boldsymbol{n}} \boldsymbol{x}_{\boldsymbol{i'}}{}^{\boldsymbol{2}}\right). \tag{13}$$

Based on that, we can get:

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \boldsymbol{x}_{\boldsymbol{i}} \cdot \left(\frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x}_{\boldsymbol{i}}\boldsymbol{y}_{\boldsymbol{i}}}{\sum_{\boldsymbol{i'=1}}^{\boldsymbol{n}} \boldsymbol{x}_{\boldsymbol{i'}}{}^{\boldsymbol{2}}}\right) \tag{14}$$

After that, we just need to place $\boldsymbol{y_i}$ outside the fraction and bring the summation sign before the fraction to get the following formula:

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \sum_{\boldsymbol{i'=1}}^{\boldsymbol{n}} \frac{\boldsymbol{x}_{\boldsymbol{i'}}\boldsymbol{x}_{\boldsymbol{i}}}{\sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{x}_{\boldsymbol{j}}{}^{\boldsymbol{2}}} \boldsymbol{y}_{\boldsymbol{i'}} \tag{15}$$

From the above equation, we can easily see that the form of this equation is similar to the following equation (16):

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \sum_{\boldsymbol{i'=1}}^{\boldsymbol{n}} \boldsymbol{a}_{\boldsymbol{i'}}\boldsymbol{y}_{\boldsymbol{i'}} \tag{16}$$

Therefore, compare with the equation (15) and (16), we can get the $\boldsymbol{a_{i'}}$,that is:

$$\boldsymbol{a}_{\boldsymbol{i'}} = \frac{\boldsymbol{x}_{\boldsymbol{i'}}\boldsymbol{x}_{\boldsymbol{i}}}{\sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{x}_{\boldsymbol{j}}{}^{\boldsymbol{2}}} \tag{17}$$

3.7.6 Solution:
For the simple linear regression we know:

$$\boldsymbol{y} = \hat{\boldsymbol{\beta}}_{\boldsymbol{0}} + \hat{\boldsymbol{\beta}}_{\boldsymbol{1}}\boldsymbol{x} \tag{18}$$

Then, from the equation (3.4) in the textbook we can get:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{19}$$

Then, we can change its form:

$$\hat{\beta}_0 + \hat{\beta}_1 \bar{x} - \bar{y} = 0 \tag{20}$$

Then, if the least squares line always passes through the point $(\bar{x}, \bar{y})$, we can easily see that the equation (18) will become:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \tag{21}$$

Then, we get the right value of $\hat{\beta}_0$ from equation (19) and substitute that into equation (21), we can get:

$$\bar{y} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} \tag{22}$$

Because the right side of the equation (22) is $\bar{y}$, and the left side of the equation (22) is also $\bar{y}$. So, the equation (22) is always right when the least squares line always passes through the point $(\bar{x}, \bar{y})$.

In a conclusion, in the case of simple linear regression, the least squares line always passes through the point $(\bar{x}, \bar{y})$.

3.7.7 Proof:
In the case of simple linear regression, R2 statistic is defined as:

$$R^2 = \frac{TSS - RSS}{TSS} \tag{23}$$

Then, TSS and RSS is defined as:

$$TSS = \sum (y_i - \bar{y})^2 \tag{24}$$

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{25}$$

Then, the correlation between X and Y is:

$$Cor(X, Y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{26}$$

Next, we start to proof the conclusion that this problem gives us. The process is shown below:

Firstly, we substitute the defintion of TSS and RSS into R^2, that is:

$$R^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{27}$$

Then, we look at the numerator and suppose the value of it is Z, by using the square variance formula it will become:

$$\begin{aligned} Z &= \sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n} \left[(y_i - \bar{y}) - (y_i - \hat{y}_i) \right] \left[(y_i - \bar{y}) + (y_i - \hat{y}_i) \right] \\ &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})(2y_i - \bar{y} - \hat{y}_i) \end{aligned} \tag{28}$$

Next,in the case of simple linear regression of Y onto X, we should know that:

$$\hat{\boldsymbol{\beta}}_{\mathbf{0}} = \bar{\boldsymbol{y}} - \hat{\boldsymbol{\beta}}_{\mathbf{1}}\bar{\boldsymbol{x}} \tag{29}$$

$$\hat{\boldsymbol{\beta}}_{\mathbf{1}} = \frac{\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}})}{\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^{\mathbf{2}}} \tag{30}$$

So, we can substitute the expression for $\hat{\boldsymbol{\beta}}_{\mathbf{0}}$ into $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$, that is:

$$\begin{aligned}\hat{\boldsymbol{y}}_{\boldsymbol{i}} &= \hat{\boldsymbol{\beta}}_{\mathbf{0}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}\boldsymbol{x_i} \\ &= \bar{\boldsymbol{y}} - \hat{\boldsymbol{\beta}}_{\mathbf{1}}\bar{\boldsymbol{x}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}\boldsymbol{x_i} \\ &= \bar{\boldsymbol{y}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}(\boldsymbol{x_i} - \bar{\boldsymbol{x}}) \end{aligned} \tag{31}$$

Therefore, we can use the above conclusion from equation (31) to simplify the partial expression of $\boldsymbol{Z}$, that is:

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} - \bar{\boldsymbol{y}} = \hat{\boldsymbol{\beta}}_{\mathbf{1}}(\boldsymbol{x_i} - \bar{\boldsymbol{x}}) \tag{32}$$

$$\begin{aligned}\mathbf{2}\boldsymbol{y_i} - \bar{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{\boldsymbol{i}} &= \mathbf{2}\boldsymbol{y_i} - \bar{\boldsymbol{y}} - \bar{\boldsymbol{y}} - \hat{\boldsymbol{\beta}}_{\mathbf{1}}(\boldsymbol{x_i} - \bar{\boldsymbol{x}}) \\ &= \mathbf{2}(\boldsymbol{y_i} - \bar{\boldsymbol{y}}) - \hat{\boldsymbol{\beta}}_{\mathbf{1}}(\boldsymbol{x_i} - \bar{\boldsymbol{x}}) \end{aligned} \tag{33}$$

After that, we can substitute the equation (32) and (33) to the expression of $\boldsymbol{Z}$, that is:

$$\begin{aligned}\boldsymbol{Z} &= \sum_{i=1}^{n}\hat{\boldsymbol{\beta}}_{\mathbf{1}}(\boldsymbol{x_i} - \hat{\boldsymbol{x}})\left[\mathbf{2}(\boldsymbol{y_i} - \bar{\boldsymbol{y}}) - \hat{\boldsymbol{\beta}}_{\mathbf{1}}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})\right] \\ &= \hat{\boldsymbol{\beta}}_{\mathbf{1}}\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})\left[\mathbf{2}(\boldsymbol{y_i} - \bar{\boldsymbol{y}}) - \hat{\boldsymbol{\beta}}_{\mathbf{1}}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})\right] \\ &= \hat{\boldsymbol{\beta}}_{\mathbf{1}}\left[\mathbf{2}\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}}) - \hat{\boldsymbol{\beta}}_{\mathbf{1}}\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^{\mathbf{2}}\right] \end{aligned} \tag{34}$$

From the equation (30), we know that the equation can be simplified as:

$$\begin{aligned}\boldsymbol{Z} &= \hat{\boldsymbol{\beta}}_{\mathbf{1}}\left[\mathbf{2}\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}}) - \hat{\boldsymbol{\beta}}_{\mathbf{1}}\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^{\mathbf{2}}\right] \\ &= \hat{\boldsymbol{\beta}}_{\mathbf{1}}\left[\mathbf{2}\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}}) - \sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}})\right] \\ &= \hat{\boldsymbol{\beta}}_{\mathbf{1}}\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}}) \\ &= \frac{\left[\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}})\right]^{\mathbf{2}}}{\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^{\mathbf{2}}} \end{aligned} \tag{35}$$

At this point, we can substitute our final reduced expression for $\boldsymbol{Z}$ into $\boldsymbol{R^2}$, that is:

$$\boldsymbol{R^2} = \frac{\left[\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{y_i} - \bar{\boldsymbol{y}})\right]^{\mathbf{2}}}{\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^{\mathbf{2}}\sum_{i=1}^{n}(\boldsymbol{y_i} - \bar{\boldsymbol{y}})^{\mathbf{2}}} \tag{36}$$

Compare with the equation (26), we can finally get:

$$\boldsymbol{R^2} = \boldsymbol{Cor^2(X, Y)} \tag{37}$$

So, in the case of simple linear regression of $\boldsymbol{Y}$ onto $\boldsymbol{X}$, the $\boldsymbol{R^2}$ statistic is equal to the square of the correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$.