DATA 558: Homework 2

Trevor Nims

April 14, 2021

Exercise 1

In this section, I will that the coefficient of determination:

$$R^2 = \frac{\text{TSS-RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

With:

$$\text{RSS} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \text{ and } \text{TSS} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Has a domain given by:

$$0 \leq R^2 \leq 1$$

I will do this by showing that the TSS is equal to the sum of the RSS and the ESS, that is:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Proof. Let us begin with the LHS:

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \bar{y})^2 &= \sum_{i=1}^{n}(y_i - \bar{y} + \hat{y}_i - \hat{y}_i)^2 \\
&= \sum_{i=1}^{n}((\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i))^2 \\
&= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})
\end{aligned}$$

Now recall that $\hat{\beta}$ is chosen to minimize the RSS, that is to say:

$$\frac{\partial}{\partial \beta^{(j)}} RSS = -2\sum_{i=1}^{n}(y_i - \beta^T x_i)x_i^{(j)}$$

And with $\hat{\beta}$ we have:

$$-2\sum_{i=1}^{n}(y_i - \beta^T x_i)x_i^{(j)} = 0$$

Now, breaking down the final addend from our initial derivation we have:

$$2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 2\Bigg(\sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i) - \bar{y}\sum_{i=1}^{n}(y_i - \hat{y}_i)\Bigg)$$

However, because we know:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$$

We just need to show:

$$\sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i) = 0$$

Consider the partial derivative of the RSS with respect to $\beta^{(j)}$, then we have:

$$\begin{aligned}
\sum_{i=1}^{n} x_i(y_i - \hat{y}_i) &= 0 \\
\sum_{i=1}^{n} \beta^T x_i(y_i - \hat{y}_i) &= 0 \\
\sum_{i=1}^{n} \hat{\beta}^T x_i(y_i - \hat{y}_i) + \sum_{i=1}^{n} \hat{\beta}_0(y_i - \hat{y}_i) &= 0 \\
\sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}^T x_i)(y_i - \hat{y}_i) &= 0 \\
\sum_{i=1}^{n} \hat{y}_i(y_i - \hat{y}_i) &= 0
\end{aligned}$$

Therefore:

$$2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$$

Giving us:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

And we are done. $\square$

Exercise 2

Now, I will prove the following identity:

$$E_y[(y - \bar{\beta}^T x)^2] = E_y[(y - f(x))^2] + (f(x) - \bar{\beta}^T x)^2$$

With $f(x)$ leearned from the data, that is:

$$f(x) \approx \hat{\beta}^T x$$

Proof. Beginning with the LHS:

$$\begin{aligned}
E_y[(y - \bar{\beta}^T x)^2] &= E_y[((y - f(x)) + (f(x) - \bar{\beta}^T x))^2] \\
&= E_y[(y - f(x))^2] + 2E_y[(y - f(x))(f(x) - \bar{\beta}^T x)] + E_y[(f(x) - \bar{\beta}^T x))^2] \\
&= E_y[(y - f(x))^2] + (f(x) - \bar{\beta}^T x))^2 + 2E_y[(y - f(x))]E_y[(f(x) - \bar{\beta}^T x)]
\end{aligned}$$

Note that we were able to distribute the expectation across the product because $(f(x) - \bar{\beta}^T x)$ is independent of y. Moreover, because the expected value of the residual is equal to zero:

$$E_y[(y - f(x))] = 0$$

The final term in the addition is eliminated, leaving us with:

$$E_y[(y - \bar{\beta}^T x)^2] = E_y[(y - f(x))^2] + (f(x) - \bar{\beta}^T x))^2$$

And we are done. $\square$