Homework 1 - Machine Learning

Felix Nguyen

February 2019

1 Exercise 1

From OLS properties, we have:

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
\hat{\beta} &= (X'X)^{-1}X'(\beta X + \epsilon) \\
\hat{\beta} &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon \\
\hat{\beta} &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon
\end{aligned}$$

Since we have $(X'X)^{-1}X'X = I$, this could be reduced to

$$\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon$$

The assumptions of OLS estimators say that $E(\epsilon) = 0$, and X and ϵ are independent, so in any case, we would have:

$$\begin{aligned}
\hat{\beta} &= \beta + (X'X)^{-1}X'\epsilon \\
E(\hat{\beta}) &= \beta + E((X'X)^{-1}X')E(\epsilon) \\
E(\hat{\beta}) &= \beta + 0 = \beta
\end{aligned}$$

as long as the assumptions are hold.

2 Exercise 2

Since 1 occurs n_1 times in your sample and 0 occurs $n - n_1$ times in the sample, we have $\bar{x} = \frac{1n_1 + 0(n-n_1)}{n} = \frac{n_1}{n}$. Keep in mind that $\sum_{i=1}^{n}(x_i) = n\bar{x}$, and similar for y, from general OLS properties, we have:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_1 - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})}{\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)} \\
&= \frac{\sum_{i=1}^{n}(x_i y_i) - n\bar{x}\bar{y} - \bar{x}n\bar{y} + n\bar{x}\bar{y}}{\sum_{i=1}^{n}(x_i^2) - 2n\bar{x}^2 + n\bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i y_i) - n\bar{x}\bar{y}}{\sum_{i=1}^{n}(x_i^2) - n\bar{x}^2}
\end{aligned}$$

Since x_i is either 0 or 1, we can easily see that $\sum_{i=1}^{n}(x_i^2) = \sum_{i=1}^{n}(x_i) = n_1$. If we call $\bar{y}_{x=1}$ and $\bar{y}_{x=0}$ as mean of y when x = 1 and x = 0 respectively, we can see that $\sum_{i=1}^{n}(x_i y_i) = n_1\bar{y}_{x=1}$. We then have:

$$\hat{\beta}_1 = \frac{n_1\bar{y}_{x=1} - n_1\bar{y}}{n_1 - n_1\frac{n_1}{n}} = \frac{n_1\bar{y}_{x=1} - n_1\frac{n_1\bar{y}_{x=1}+(n-n_1)\bar{y}_{x=0}}{n}}{n_1 - n_1\frac{n_1}{n}}$$
$$= \frac{(\frac{n-n_1}{n})(\bar{y}_{x=1} - \bar{y}_{x=0})}{(\frac{n-n_1}{n})} = \bar{y}_{x=1} - \bar{y}_{x=0}$$

We have

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = \frac{n_1\bar{y}_{x=1} + (n-n_1)\bar{y}_{x=0}}{n} - (\bar{y}_{x=1} - \bar{y}_{x=0})\frac{n_1}{n}$$
$$= (\bar{y}_{x=1} - \bar{y}_{x=0})\frac{n_1}{n} - (\bar{y}_{x=1} - \bar{y}_{x=0})\frac{n_1}{n} + \bar{y}_{x=0} = \bar{y}_{x=0}$$

3 Exercise 3

We have:

$$Correl[y,\hat{y}] = \frac{Cov(y,\hat{y})}{\sqrt{var(y)var(\hat{y})}}$$
$$Thus,$$
$$Correl[y,\hat{y}]^2 = \frac{Cov(y,\hat{y})^2}{var(y)var(\hat{y})}$$

We also have $y = \hat{y} + \epsilon$, so:

$$Correl[y,\hat{y}]^2 = \frac{Cov(\hat{y}+\epsilon,\hat{y})^2}{var(y)var(\hat{y})} = \frac{(Cov(\hat{y},\hat{y}) + (Cov(\hat{y},\epsilon))^2}{var(y)var(\hat{y})}$$

By OLS assumptions, $Cov(\hat{y},\epsilon) = 0$, so:

$$Correl[y,\hat{y}]^2 = \frac{Cov(\hat{y},\hat{y})^2}{var(y)var(\hat{y})} = \frac{var(\hat{y})^2}{var(y)var(\hat{y})} = \frac{var(\hat{y})}{var(y)}$$
$$= \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}}_i)}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)}$$

We have $\bar{\hat{y}}_i = \bar{y}_i$, so,

$$Correl[y,\hat{y}]^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)} = ESS/TSS = R^2$$

4 Exercise 4

We can transform the given function to Matrix form as:

$$RSS(\beta,\lambda) := (y - X\beta')(y - X\beta) + \lambda\beta\beta'$$

Taking the first derivative for minimization, we have:

$$\begin{aligned}\frac{\delta}{\delta\beta}RSS(\beta,\lambda) := 2X'X\beta - 2X'y + 2\beta\lambda = 0 \\ 2(X'X)\beta + 2\lambda\beta = 2X'y \\ X'X\beta + I\lambda\beta = X'y\end{aligned}$$

Thus, we can derive $\beta^{r\hat{i}dge}$ as $\beta^{r\hat{i}dge} = (X'X + \lambda I)^{-1}X'y$.