KAZAKH-BRITISH TECHNICAL UNIVERSITY

Introduction to Machine Learning
Week 5

Olivier JAYLET

School of Information Technology and Engineering

Ordinary least square (OLS)

From the model :

$$y_i = \alpha + \beta x_i + \epsilon_i \quad (1)$$

- y_i : is known
- x_i : is known
- $\epsilon_i = y_i - \alpha - \beta x_i$
- α & β are unknown parameters (intercept & slope)

The purpose of linear regression is to estimate $\hat{\alpha}$ & $\hat{\beta}$ such that the SSR is minimized.

Minimization problem

The minimization problem can be written as :

$$\min_{\alpha,\beta} SSR \tag{2}$$

$$\min_{\alpha,\beta} \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2. \tag{3}$$

There exists two type of solutions. Using iterative algorithm, or finding the analytical solution.

OLS estimator

$$\min_{\alpha,\beta} \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \tag{4}$$

$$\frac{\partial SSR}{\partial \hat{\alpha}} = 0 \tag{5}$$

$$\frac{\partial SSR}{\partial \hat{\beta}} = 0 \tag{6}$$

OLS estimator

By derivating and solving for both parameters, you should find :

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{7}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{8}$$

Multi-linear regression

So far, we solve the linear equation for the uni-variate case, i.e. we estimated the relation in between one explanatory variable and one target variable.
What if we want to estimate y with more than one x ?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{9}$$

Matrix form

We can redefine the linear regression using matrices :

$$y = X.\beta + \epsilon \tag{10}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} . \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{11}$$

Residuals

As for SLR, we can compute the residuals:

$$y = X\beta + \epsilon \tag{12}$$

$$\epsilon = y - X\hat{\beta} \tag{13}$$

SSR

For the SLR we had :

$$SSR = \sum_{i=1}^{n} \epsilon_i^2 \tag{14}$$

In MLR, the residuals are also in a vector (one residual per observation/prediction). But we can re write it in matrix form as :

$$\epsilon^T.\epsilon = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_n \end{bmatrix} . \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{15}$$

$$\epsilon^T.\epsilon = \begin{bmatrix} \epsilon_1.\epsilon_1 & \epsilon_2.\epsilon_2 & \cdots & \epsilon_n.\epsilon_n \end{bmatrix} \tag{16}$$

SSR

Looking at the residuals, and the vector form of SSR, we can develop the SSR :

$$\begin{aligned}
\epsilon'\epsilon &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= y'y - \hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\
&= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}
\end{aligned}$$

Minimization problem for MLR

$$\min_{\hat{\beta}} \sum_{i=1}^{N} \epsilon_i^2 \tag{17}$$

$$\min_{\hat{\beta}} \epsilon' \epsilon \tag{18}$$

$$\min_{\hat{\beta}} y'y - 2\hat{\beta}' X' y + \hat{\beta}' X' X \hat{\beta} \tag{19}$$

OLS

By computing the FOC and solving for $\hat{\beta}$, you should find :

$$\hat{\beta} = (X'X)^{-1}X'y \tag{20}$$