COMP 5360 / MATH 4100: Introduction to Data Science
Notes on Simple Linear Regression
February 5, 2024

Data. We have n samples $(x_1, y_1),\ (x_2, y_2), \ldots, (x_n, y_n)$.

Model. $y \sim \beta_0 + \beta_1 x$

Goal. Find the best values of β_0 and β_1, denoted $\hat{\beta}_0$ and $\hat{\beta}_1$, so that the prediction $y = \hat{\beta}_0 + \hat{\beta}_1 x$ "best fits" the data.

Theorem. The best parameters in the *least squares sense* are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} \qquad \text{and} \qquad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x},$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$ and $\overline{y} = \frac{1}{n}\sum_{i=1}^n y_i$.

Least squares sense. Consider the errors, defined

$$e_i = y_i - \beta_0 - \beta_1 x_i.$$

The idea then is to minimize the total squared error over all β_0 and β_1. We define the total squared error,

$$J(\beta_0, \beta_1) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2.$$

How do we minimize J with respect to β_0 and β_1? Set the partial derivatives to zero and solve for the minimum values. Before we minimize J, let's do a couple of exercises to recall how we solve optimization problems like this in general.

Exercise 1. Minimize $f(x) = (x-2)^2$.

Exercise 2. Minimize $f(x, y) = x^2 + y^2$.

Proof of theorem. We first set the partial derivative of $J(\beta_0, \beta_1)$ with respect to β_0 to 0 and evaluate at $(\hat{\beta}_0, \hat{\beta}_1)$ to obtain

$$\begin{aligned}
\frac{\partial J}{\partial \beta_0}(\hat{\beta}_0, \hat{\beta}_1) &= \sum_{i=1}^n 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0 \\
\implies \sum y_i - \sum \hat{\beta}_0 - \sum \hat{\beta}_1 x_i &= 0 \\
\implies \sum y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum x_i &= 0 \\
\implies \frac{1}{n}\sum y_i - \hat{\beta}_0 - \hat{\beta}_1 \frac{1}{n}\sum x_i &= 0.
\end{aligned}$$

Now, using the definition of $\overline{x}$ and $\overline{y}$, we have

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}. \tag{1}$$

So, if we know $\hat{\beta}_1$, then we can determine $\hat{\beta}_0$ from (1).

To determine $\hat{\beta}_1$, we set the partial derivative of $J(\beta_0, \beta_1)$ with respect to β_1 to 0 and evaluate at $(\hat{\beta}_0, \hat{\beta}_1)$ to obtain

$$\begin{aligned} \frac{\partial J}{\partial \beta_1}(\hat{\beta}_0, \hat{\beta}_1) &= \sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) = 0 \\ \implies \sum x_i y_i - \hat{\beta}_0 \sum x_i - \hat{\beta}_1 \sum x_i^2 &= 0. \end{aligned}$$

Using (1) and the definition of $\overline{x}$, we have

$$\begin{aligned} \sum x_i y_i - (\overline{y} - \hat{\beta}_1 \overline{x}) n \overline{x} - \hat{\beta}_1 \sum x_i^2 &= 0 \\ \implies \left(\sum x_i^2 - n\overline{x}^2 \right) \hat{\beta}_1 &= \sum x_i y_i - n\overline{xy} \end{aligned}$$

This gives

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\overline{xy}}{\sum x_i^2 - n\overline{x}^2}. \tag{2}$$

We now just have to manipulate the numerator and denominator in (2) to agree with the statement in the theorem. We compute

$$\begin{aligned} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) &= \sum x_i y_i - \sum \overline{x} y_i - \sum x_i \overline{y} + \sum \overline{xy} \\ &= \sum x_i y_i - n\overline{xy} - n\overline{xy} + n\overline{xy} \\ &= \sum x_i y_i - n\overline{xy}, \end{aligned}$$

so the numerators agree. To see that $\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum x_i^2 - n\overline{x}^2$ in the denominators, we just set $x_i = y_i$ in the above calculation. $\square$