Linear regression notes

John Bogovic

2019 July

1 Simple linear regression (fit a 1d line)

1.1 The problem

We are given the value of a variable, x, we would like to predict the value of another variable y. We have many pairs of examples:

$$\begin{aligned} & x_1, y_1 \\ & x_2, y_2 \\ & \vdots \\ & x_N, y_N \end{aligned} \tag{1}$$

Linear regression (1d) does this by finding the linear function that gives the best predictions. The functions we have to consider are:

$$\hat{y} = ax + b \tag{2}$$

Where we wrote $\hat{y}$ instead of y to indicate that it is an estimate, or prediction, and not the true value of y for the given x. Another way to think of the task is that we need to find the values a and b that give us the best results. The values a and b are called "parameters" of the function. How do we measure how good the predictions are?

1.2 "Cost function" - measuring goodness of the prediction

The most common way is to use the "sum of squared differences" (SSD) also called "residual sum of squares." SSD is computed like this:

$$SSD(a, b) = \sum_i (y_i - (ax_i + b))^2. \tag{3}$$

Notice that $ax_i + b$ is the value of y predicted by our function for the input x_i.

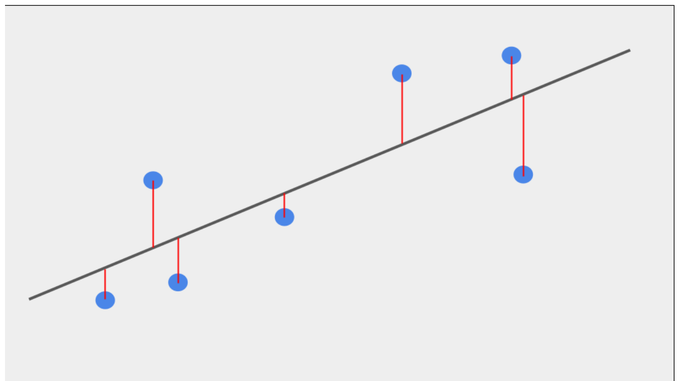

Figure 1: Adding the (squared) lengths of the red lines will tell us how good this fit is.

1.3 Rewrite the problem using linear algebra

It might seem strange to do this now, but it will help us find a solution to the problem and help us use the technique when we have many input and/or output variables.

First, let's stack the y_i in a vector.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \tag{4}$$

Next, convince yourself that if we can write a vector of function predictions like this:

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{5}$$