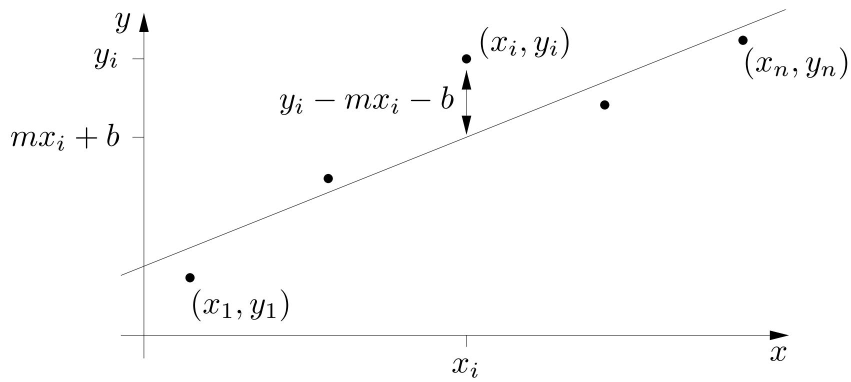
Linear Regression

Imagine an experiment in which you measure one quantity, call it y, as a function of a second quantity, say x. For example, y could be the current that flows through a resistor when a voltage x is applied to it. Suppose that you measure n data points $(x_1, y_1), \cdots, (x_n, y_n)$ and that you wish to find the straight line $y = mx + b$ that fits the data best. If the data point



(x_i, y_i) were to land exactly on the line $y = mx + b$ we would have $y_i = mx_i + b$. If it doesn't land exactly on the line, the vertical distance between (x_i, y_i) and the line $y = mx + b$ is $|y_i - mx_i - b|$. That is, the discrepancy between the measured value of y_i and the corresponding idealized value on the line is $|y_i - mx_i - b|$. One measure of the total discrepancy for all data points is $\sum_{i=1}^{n} |y_i - mx_i - b|$. A more convenient measure, which avoids the absolute value signs, is

$$D(m, b) = \sum_{i=1}^{n} (y_i - mx_i - b)^2$$

We will now find the values of m and b that give the minimum value of $D(m, b)$. The corresponding line $y = mx + b$ is generally viewed as the line that fits the data best.

You learned in your first Calculus course that the value of m that gives the minimum value of a function of one variable $f(m)$ obeys $f'(m) = 0$. The analogous statement for functions of two variables is the following. First pretend that b is just a constant and compute the derivative of $D(m, b)$ with respect to m. This is called the partial derivative of $D(m, b)$ with respect to m and denoted $\frac{\partial D}{\partial m}(m, b)$. Next pretend that m is just a constant and compute the derivative of $D(m, b)$ with respect to b. This is called the partial derivative of $D(m, b)$ with respect to b and denoted $\frac{\partial D}{\partial b}(m, b)$. If (m, b) gives the minimum value of $D(m, b)$, then

$$\tfrac{\partial D}{\partial m}(m, b) = \tfrac{\partial D}{\partial b}(m, b) = 0$$

For our specific $D(m, b)$

$$\tfrac{\partial D}{\partial m}(m, b) = \sum_{i=1}^{n} 2(y_i - mx_i - b)(-x_i)$$

$$\tfrac{\partial D}{\partial b}(m, b) = \sum_{i=1}^{n} 2(y_i - mx_i - b)(-1)$$

It is important to remember here that all of the x_i's and y_i's here are given numbers. The only unknowns are m and b. The two partials are of the forms

$$\begin{aligned} \tfrac{\partial D}{\partial m}(m,b) &= 2c_{xx}m + 2c_x b - 2c_{xy} \\ \tfrac{\partial D}{\partial b}(m,b) &= 2c_x m + 2nb - 2c_y \end{aligned}$$

where the various c's are just given numbers whose values are

$$c_{xx} = \sum_{i=1}^{n} x_i^2 \qquad c_x = \sum_{i=1}^{n} x_i \qquad c_{xy} = \sum_{i=1}^{n} x_i y_i \qquad c_y = \sum_{i=1}^{n} y_i$$

So the value of (m, b) that gives the minimum value of $D(m, b)$ is determined by

$$\begin{aligned} c_{xx}m + c_x b &= c_{xy} \qquad (1) \\ c_x m + nb &= c_y \qquad (2) \end{aligned}$$

This is a system of two linear equations in the two unknowns m and b, which is easy to solve:

$$\begin{aligned} n(1) - c_x(2): &\qquad [nc_{xx} - c_x^2]m = nc_{xy} - c_x c_y &\qquad \Longrightarrow \qquad m = \frac{nc_{xy} - c_x c_y}{nc_{xx} - c_x^2} \\ c_x(1) - c_{xx}(2): &\qquad [c_x^2 - nc_{xx}]b = c_x c_{xy} - c_{xx} c_y &\qquad \Longrightarrow \qquad b = \frac{c_{xx} c_y - c_x c_{xy}}{nc_{xx} - c_x^2} \end{aligned}$$