Least Squares Regression

September 7, 2023

Consider a set of points (x_i, y_i) for $i = 1, 2, 3, \ldots, N$. The problem of least squares regression is one that will help us try to discern the relationship between $\{x_i, i = 1, 2, 3, \ldots, N\}$ and $\{y_i, i = 1, 2, 3, \ldots, N\}$. At this point, let's suppose that x_i and y_i are just numbers, but we'll need to generalize to the case of vectors very soon.

Suppose I have some function $f(x, a)$. I want to pick a function so that the qualitative behavior of f changes significantly with different values of a[1]. The least squares problem is to find the value of a that minimizes

$$S = \sum_{i=1}^{N} (f(x_i, a) - y_i)^2. \tag{1}$$

Let's suppose that $f(x_i, a) = a$. Then

$$S = \sum_{i=1}^{N} (a - y_i)^2. \tag{2}$$

To minimize this with respect to a, we take

$$0 = \frac{\partial S}{\partial a} = \sum_{i=1}^{N} (a - y_i). \tag{3}$$

Rearranging terms,

$$Na = \sum_{i=1}^{N} y_i. \tag{4}$$

or rather

$$a = \bar{y} \tag{5}$$

This is one of the special properties of the arithmetic mean; it minimizes least squared distance.

1 Linear Least Squares

Let's do another example where, now, we'll treat $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}^m$ as vectors and we will use the fit function

$$f(x) = Ax \tag{6}$$

where A is a matrix. You might wonder why there is no affine term, but, as we will see, it is straightforward to extend to that case from this one.

Now, with

$$X^{\mathrm{T}} = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_N \end{pmatrix} \tag{7}$$

and

$$Y^{\mathrm{T}} = \begin{pmatrix} y_1 & y_2 & y_3 & \ldots & y_N \end{pmatrix} \tag{8}$$

[1] a may be a vector, for example.

$$
\begin{aligned}
S &= \sum_{i=1}^{N} (Ax_i - y_i)^2 = \|AX - Y\|_F^2 = \mathrm{Tr}[(AX - Y)^{\mathrm{T}}(AX - Y)] \\
&= \mathrm{Tr}[X^{\mathrm{T}}A^{\mathrm{T}}AX - X^{\mathrm{T}}A^{\mathrm{T}}Y - Y^{\mathrm{T}}AX + Y^{\mathrm{T}}Y].
\end{aligned} \tag{9}
$$

All we have to do at this point is to take derivatives

2 Not done yet