Analytical solution

Welcome back!

Now that we introduced the concept of least squares, let's try to translate this concept into a mathematical form and find an analytical solution for it.

Coming back to our line fitting problem, we are trying to determine what a good solution would be for this inconsistent system of equations.

Say we have a solution xhat.

The least squares principle states that the best solution for the unknown parameters in the vector x is the one which results in the set of new "adjusted" observations yhat which is the closest or has a minimum distance to the original observation vector y.

In a mathematical form, this condition of "closeness" or "minimum distance" for a particular estimate xhat is defined as the length of the vector ehat.

This vector is the remaining part of the vector y with respect to yhat, and so in the rest of this course we call it: the vector of "residuals".

So, based on the least squares principle, the least squares estimate should minimises the length of the error vector, or the difference between observation vector y and Ax.

The length of the error vector or the norm of the vector e is defined as the square root of the summation over squared elements of the vector.

Note that, the summation under square root can be written in the vector form as (e transpose e).

The LS estimate of the unknowns parameters can then be found by the minimisation of the norm or length of the vector e.

The minimization of the norm of a vector, is equivalent with the minimisation of its squared norm.

So we can compute the least squares solution by minimising the squared norm of the vector e.

Recall that, the vector of residuals or error vector is the difference between the observation vector y and the adjusted observation vector which can be computed by A x.

By inserting (y-Ax) instead of (e), the minimisation problem can be written as the minimising the squared norm of (y-Ax).

Therefore, as we can see, xhat of least squares minimises the sum of the squares of the error vector e.

This is actually the main idea behind the name “least squares”.

We know from calculus that in order to find the solution to this minimization problem, we should take the first derivative from this multivariate objective function and set it to zero.

From calculus, this first derivative can be derived as the following.

By the way, if you are interested in the detailed mathematical derivation of this derivative, please look at the reading materials of this week.

Now by setting this derivative to zero, we get: A transpose Ax equals A transpose y.

This new equation is called “normal equation”.

The solution to the normal equation provides us the least squares estimate.

The solution can simply computed by taking the inverse of the left hand side of the normal equation and multiplying it by its right hand side.

The least squares estimate is then computed as: Inverse of A transpose A multiplied by A transpose y.

Therefore, if we have an observation vector y and a design matrix A, we can simply put them in this simple equation and compute the least squares estimate of x.

To summarise, for any given estimation problem where we relate our unknowns parameters x to observations y, using a functional model described by the design matrix A, the least squares estimate can be computed by this simple equation.

In addition to (xhat), the least squares estimate of observations (yhat) can be also computed by inserting (xhat) into the functional model.

And finally the least squares estimates of residuals (ehat) can be also computed easily as a difference between the original observations in the vector (y) and the adjusted observations (yhat).

Now that we know all about the mathematics of least squares, let’s look at its application to our line fitting problem, and also at some more exercises.