Least Square Regression

Li Ying Yin Simon

April 2021

1 Least Square Regression

Here explains the least square regression. First we start with a system of linear equation. Consider the following equations

$$\begin{cases} x = 6 \\ x + y = 0 \end{cases} \tag{1}$$

or equivalently a 2×2 system of linear equation.

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix} \tag{2}$$

The answer for the above system is $x = 6, y = -6$. We can view the problem as finding the intersection point of two straight lines. But also we can view it as given two input points $(0, 6), (1, 0)$ and find a line which passes through this two points.
The first point $(0, 6)$ is on the line $y = mx + c$ if $6 = c + m * 0$.
The second point $(1, 0)$ is on the line $y = mx + c$ if $0 = c + m * 1$.
m is the slope and c is the y-intercept. We have the following system of equations, which is same as 2

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c \\ m \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix} \tag{3}$$

The above system is solvable. However, if we have one extra point $(2, 0)$. The system of linear equation becomes

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} \tag{4}$$

4 is not solvable since we cannot find a straight line passes through these three points. However, we could try to find a closest line to these three points.
we have generally $m \times n$ system of linear equations. Here we consider $m \geq n$ i.e. more input points than dimension of the points. The system of linear equation could be written as $Ax = b$.

$$A = \begin{pmatrix} A_{11} & A_{12} & ... & A_{1n} \\ A_{21} & A_{22} & ... & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & ... & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ ... \\ b_n \end{pmatrix} \tag{5}$$

1.1 Derivation

We are trying to solve this problem by optimization method and by linear algebra method.

1.1.1 Optimization Method

Since we are trying to find the best approximation, we can introduce a squared error $E = ||Ax - b||^2$. We formulate the problem as an minimization problem:

$$E = min\,||Ax - b||^2, \tag{6}$$

We want to find $\nabla f(x) = 0$.
Let $f(x) = ||Ax - b||^2$.

$$\begin{aligned} f(x) &= ||Ax - b||^2 \\ &= (Ax - b)^T(Ax - b) \\ &= (Ax)^T Ax - (Ax)^T b - b^T Ax + b^T b \\ &= x^T A^T Ax - x^T A^T b - b^T Ax + b^T b \end{aligned}$$

$$\begin{aligned} f(x + td) &= (x + td)^T A^T A(x + td) - (x + td)^T A^T b - b^T A(x + td) + b^T b \\ \frac{\partial f(x + td)}{\partial t} &= d^T A^T A(x + td) + (x + td)^T A^T Ad - d^T A^T b - b^T Ad \end{aligned}$$

$$\begin{aligned} \left.\frac{\partial f(x + td)}{\partial t}\right|_{t=0} &= d^T A^T Ax + x^T A^T Ad - d^T A^T b - b^T Ad \\ &= (d^T A^T Ax)^T + x^T A^T Ad - (d^T A^T b)^T - b^T Ad \quad \text{(We can take transpose since it is a scalar)} \\ &= x^T A^T Ad + x^T A^T Ad - b^T Ad - b^T d \\ &= 2((x^T a^T A - b^T A)^T, d) \\ &= 2(A^T Ax - A^T b, d) \\ &= 2(A^T(Ax - b), d) \\ &= f'(x)d \\ &= (\nabla f(x), d) \end{aligned}$$

Thus,

$$\nabla f(x) = 2A^T(Ax - b) = 0$$

$$A^T Ax = A^T b \tag{7}$$

$$x = (A^T A)^{-1} A^T b, \tag{8}$$

given that $A^T A$ has full rank i.e. invertible.

1.1.2 Linear Algebra Method

We can split b into two parts.

$$Ax = b = p + e, \tag{9}$$

p(projection) is in the column space which is the space span by the column vectors of A.
The perpendicular part e(error) which is in the nullspace of A^T i.e. e is in $N(A^T)$.
(since column space $\perp N(A^T)$). Thus,

$$A^T(e) = 0 \tag{10}$$

$$A^T(Ax - b) = 0 \tag{11}$$

$$x = (A^T A)^{-1} A^T b, \tag{12}$$

given that $A^T A$ has full rank i.e. invertible.

Geometrically, p is the projection of b onto the column space of A.

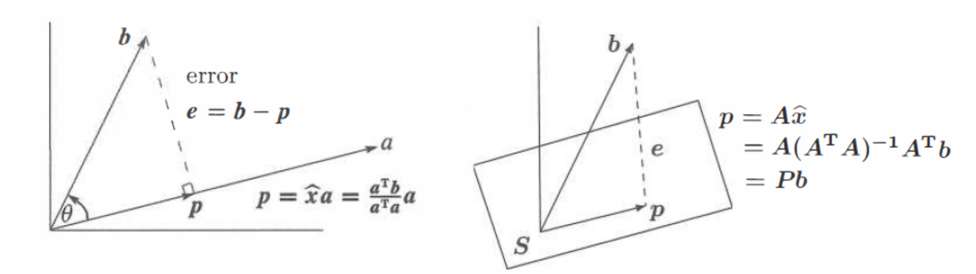

Figure 1: left: p is the projection of b onto a line a, right: p is the projection of b onto a S which is the column space of A[1].

References

[1] Gilbert Strang, Gilbert Strang, Gilbert Strang, and Gilbert Strang. *Introduction to linear algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA, 1993.