Wiki: https://en.wikipedia.org/wiki/Least_squares

- When the observations come from an exponential family and mild conditions are satisfied, least-squares estimates and maximum-likelihood estimates are identical.[1] The method of least squares can also be derived as a method of moments estimator.

Wiki: https://en.wikipedia.org/wiki/Linear_least_squares_(mathematics)#Derivation_of_the_normal_equations

1 Least Square

$$\boldsymbol{Ax} = \boldsymbol{b} \tag{1-1}$$

Where matrix A is $m \times n$, x is one $n \times 1$ vector, and b is $m \times 1$ vector

1.1 Linear Least Square

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^{\boldsymbol{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\boldsymbol{H}}\boldsymbol{b} \tag{1-2}$$

By rewriting the equation (1-2) to

$$\begin{gathered} \boldsymbol{A}^{\boldsymbol{H}}\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^{\boldsymbol{H}}\boldsymbol{b} \\ \ll=\gg \boldsymbol{A}^{\boldsymbol{H}}(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) = \boldsymbol{0} \\ \ll=\gg \boldsymbol{A}^{\boldsymbol{H}}\boldsymbol{r} = \boldsymbol{0} \end{gathered} \tag{1-3}$$

The equation (1-3) indicates the error (of $\boldsymbol{b}$) or redidual vector $\boldsymbol{r}$ is **perpendicular** (**orthogornal**) to each row in the matrix $\boldsymbol{A}^{\boldsymbol{H}}$. And these rows in matrix are called **fitting functions**. Therefore, the residual, after it has been minimized, is perpendicular to the fitting functions.

And the fuction

$$\boldsymbol{A}^{\boldsymbol{H}}\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^{\boldsymbol{H}}\boldsymbol{b} \tag{1-4}$$

are called **Normal Functions** (of the linear Least Square), where the **Normal** means **perpendicular** (**orthogornal**).

2 QRD, Cholesky decomposition, SVD, ... methods

QRD, Choleky decomposition, SVD and some other analogous methods can also be used to slove the equation (1-1), however, these methods don't have this **Normal** (**perpendicular**) form, so they are not normal equations.

e.g. use QRD, the equation (1-1) can be rewrite by

$$\boldsymbol{QRx} = \boldsymbol{b} \tag{2-1}$$

One can solve equation (2-1) by using backward-substitution to the transformation function

$$\boldsymbol{Rx} = \boldsymbol{Q}^{\boldsymbol{H}}\boldsymbol{b} \tag{2-2}$$