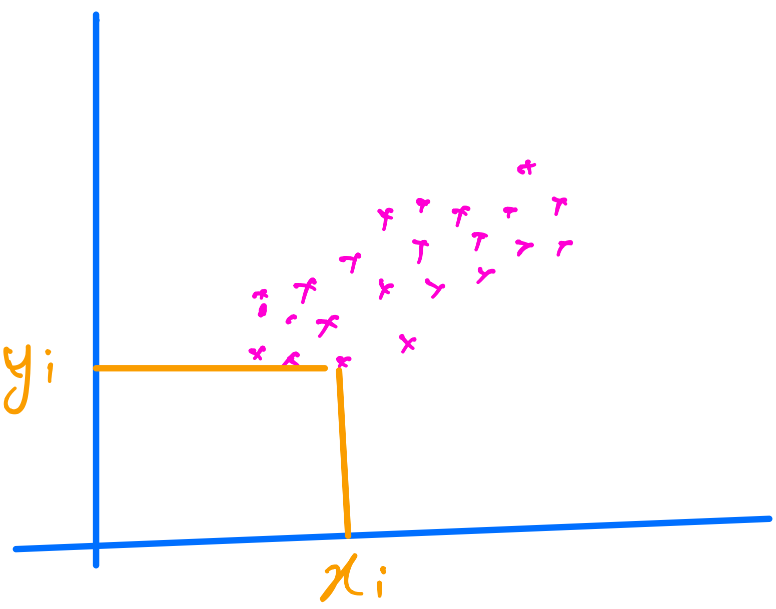

Given a set of observations y_i, what is the best model that fits the observations.

Model: $\hat{y}_i = f(x_i)$. $\longrightarrow (1)$.

where $\hat{y}_i$ is the prediction.

Then $y_i - \hat{y}_i$ is the residual or.

$y_i - (f(x_i))$ is the residual $e_i \longrightarrow (2)$

In general f is a non linear function.

but if f can be written as.

$$f = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n$$ then

f is obviously linear in the $(n-1)$ dimensional vector x_i

$$x_i = [x_{1i} \quad x_{2i} \quad \dots \quad x_{(n-1)i}] \longrightarrow (3).$$

Then $\hat{y}_i = \begin{bmatrix} x_{1i} & x_{2i} & \dots & x_{(n-1)i} & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}. \longrightarrow (4).$

or $$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_m \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{2,1} & \dots & x_{n-1,1} & 1 \\ x_{12} & x_{22} & \dots & x_{n-1,2} & 1 \\ \vdots & & & & \\ x_{1m} & x_{2m} & \dots & x_{n-1,m} & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \longrightarrow (5)$$

$$\hat{Y}_{m\times 1} = X_{m\times n} \beta_{n\times 1} \longrightarrow (6)$$

Minimize $[Y - \hat{Y}]^T [Y - \hat{Y}]$

or $[Y - X\beta]^T [Y - X\beta] \longrightarrow (7).$

or $[Y^T - \beta^T X^T][Y - X\beta]$

$$L = [Y^T Y - \beta^T X^T Y - Y^T X\beta + \beta^T X^T X\beta] \longrightarrow (8),$$

The loss L is minimized when

$\frac{\partial L}{\partial \beta}$ goes to zero

$$\frac{\partial L}{\partial \beta} = -X^T y - y^T X + 2X^T X\beta \longrightarrow (9)$$

(from Matrix Calculus)

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow -2X^T y + 2X^T X\beta = 0.$$

or $X^T X \beta = X^T y$

or $$\boxed{\beta = [X^T X]^{-1} X^T y \longrightarrow (10)}$$

is the pseudo inverse solution

→ When $m > n$ the system is overdetermined.
Due to noise the system is also inconsistent. Hence the least squares solution is the optimal estimate of the model parameters.

→ When $m = n$ the regular inverse solution exists if the equation are consistent, i.e. $\beta = X^{-1}Y$. However if the model is not consistent with the data the least squares solution works.

→ When $m < n$ the system is underdetermined. Many solutions are possible. The minimum norm solution by the method of Lagrange Multipliers is formulated as.

$$\min \beta^T \beta \longrightarrow (11).$$

$$\text{s.t} \quad Y = X\beta \longrightarrow (12).$$

The Lagrange Multiplier gives

$$\text{Min} \quad \beta^T \beta + \lambda (Y - X\beta) \longrightarrow (13).$$

which gives the pseudo inverse solution:

$$\boxed{\beta = X^T (X X^T)^{-1} Y. \longrightarrow (14)}$$

(14) is different from (10).

The Non Linear Least Squares:

$Y_{m\times 1} = f(X_{m\times n}, \beta)$. or $Y_i = f(X_i, \beta) \longrightarrow (15)$

Minimize $[Y - f(X,\beta)]^T [Y - f(X,\beta)]. \longrightarrow (16)$

$f(X,\beta) =. f(X,\beta_0) + J\delta\beta \longrightarrow (17)$

The Taylor Series Approximation.

$$J = \left.\frac{df}{d\beta}\right|_{\beta=\beta_0} \longrightarrow (18)$$

(16) becomes Minimize $[Y - f(X,\beta_0) - J\delta\beta]^T$
$[Y - f(X,\beta_0) - J\delta\beta]$

Let $Y - f(X,\beta_0)$ be Y_n. or Y_{new}

Then. Minimize $[Y_n - J\delta\beta]^T [Y_n - J\delta\beta] \longrightarrow (19)$

Then by the method of L.S

$$\boxed{\delta\beta = [J^T J]^{-1} J^T Y_n. \longrightarrow (20)}$$

Then $\beta_{n+1} = \beta_n - \delta\beta \longrightarrow (21)$.

Linearize about the new β_{n+1} till such time $\|\beta_{n+1} - \beta_n\| < \epsilon$.

The LM Algorithm:

$$\boxed{\delta\beta = \left[J^T J + \lambda I\right]^{-1} J^T y_n.}$$

λ - updates: Look at Ananth Ranganathan's material on LM.