Regression

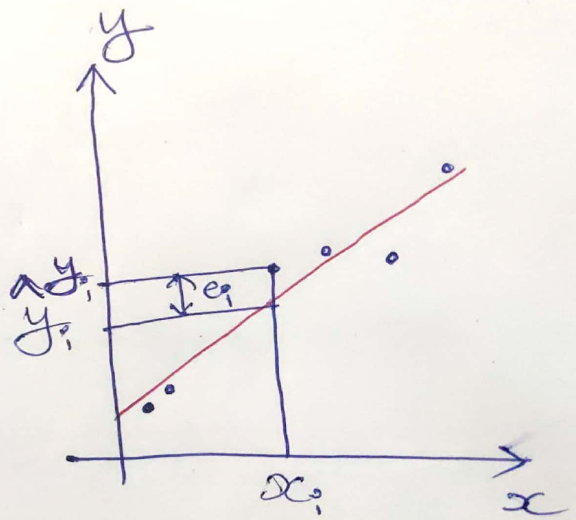
DATA $(x_1, y_1)\ (x_2, y_2) \dots (x_N, y_N)$

MODEL $[y_0 = a_0 + a_1 x]$

MODEL PREDICTION $\hat{y}_i = a_0 + a_1 x_i$

$$\text{Error} = y_i - \hat{y}_i$$

$$e_i = y_i - a_0 - a_1 x_i$$



Minimize Error in some form

Linear Regression: obtain a straight line that best fits the data.

(2)

Minimize the sum of squares of error,

Find a_0, a_1 such that

$$\min_{a_0, a_1} \sum_{i=1}^{N} (e_i)^2$$

$$\min_{a_0, a_1} \left[\sum_{i=1}^{N} (y_i - a_0 - a_1 x_i)^2 \right.$$

$$\Downarrow$$

$$\min_{a_0, a_1} \sum_{i=1}^{N} y_i^2 + a_0^2 + a_1^2 x_i^2 - 2 y_i a_0 - 2 y_i a_1 x_i + 2 a_0 a_1 x_i$$

$$\boxed{\text{Equivalent problem : Find } x \quad \min_{x} g(x) \Rightarrow \frac{dg}{dx} = 0}$$

$$\therefore \min_{a_0, a_1} S_E \Rightarrow \frac{\partial S_E}{\partial a_0} = 0 , \frac{\partial S_E}{\partial a_1} = 0$$

$$\frac{\partial S_E}{\partial a_0} = 0 \Rightarrow \sum_{i=1}^{N} \left[2a_0 - 2y_i + 2a_1 x_i \right] = 0$$

$$a_0 N - \sum_{i=1}^{N} y_i + a_1 \sum x_i = 0$$

$$a_0 N + a_1 \sum x_i = \sum_{i=1}^{N} y_i \Big\} \times \sum x_i$$

$$\frac{\partial S_E}{\partial a_1} = 0 \Rightarrow \sum_{i=1}^{N} \left[2a_1 x_i^2 - 2y_i x_i + 2a_0 x_i \right] = 0$$

$$a_0 \sum_{i=1}^{N} x_i + a_1 \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i \Big\} \times N$$

$$a_0 = y_{avg} - a_1 x_{avg} = \frac{\sum y_i}{N} - a_1 \frac{\sum x_i}{N}$$

$$a_1 = \frac{\left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right) - N \sum_{i=1}^{N} x_i y_i}{\left(\sum_{i=1}^{N} x_i\right)^2 - N \sum_{i=1}^{N} (x_i^2)}$$

Linear Least Squares

Fit a straight line $y = a_0 + a_1 x$ to the data :

x : ---- ($i=1$ to N)

y : ---- ($i=1$ to N)

Parameters a_0 & a_1 satisfy the following equations :

$$a_0 N + a_1 \sum_i x_i = \sum y_i$$

$$a_0 \sum_i x_i + a_1 \sum_i x_i^2 = \sum_i x_i y_i$$

$$\Downarrow$$

$$\begin{bmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}$$

Linear Regression for Multiple parameters

Data : $(x_1, u_1, w_1; y_1), (x_2, u_2, w_2; y_2) \cdots, (x_N, u_N, w_N; y_N)$

Model to fit : $y = a_0 + a_1 x + a_2 u + a_3 w$

$$\begin{bmatrix} 1 & x_1 & u_1 & w_1 \\ 1 & x_2 & u_2 & w_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_N & u_N & w_N \end{bmatrix} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}}_{\phi} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \Rightarrow \phi = (X^T X)^{-1} X^T Y$$

Least squares sol^n

MATLAB lsqcurvefit

Syntax : phi = lsqcurvefit(@(p,xdata) fname(p,xdata), p0, xData, yData);

phi → parameter vector , p0 → Vector of initial guesses

xData, yData → data arrays with N rows.

fname → provides $y_{model} = f(x; \phi)$