

24/03

Linear Regression

Input: A dataset with one independent variable ( $x$ ) and one dependent variable ( $y$ ).

Initialization: set the coefficients

$m \rightarrow \text{slope} \approx 0$

$b \rightarrow \text{intercept} \approx 0$

Training & Execution

For each iteration:

(\*) The best fit line is represented as  

$$y = \beta_0 + \beta_1 x + \epsilon$$

(\*) Let data points be  $(x_1, y_1) \dots (x_n, y_n)$   
 Represent the data points in matrix form

$$Y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \quad B = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad X = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

where  $\epsilon$  is the error.

$y = mx + c$  can be represented as.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 + x_1 \\ 1 + x_2 \\ \vdots \\ 1 + x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \epsilon_1$$

Then  $\beta_0$  and  $\beta_1$  can be determined by

$$\beta = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

$\beta_0$  and  $\beta_1$  values can be used to plot the best fit line and can be used to predict future values.

### Multiple Linear Regression

In multiple linear regression best fit line  $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + \epsilon$

Initializing  $b_0, b_1, \dots$  to 0

Let data points be  $(x_1, x_2, \dots, x_n, y_1)$   
 $\dots$   
 $(x_1, x_2, \dots, x_n, y_m)$  where  $x_i \forall i \in \{0, n\}$   
 represent independent variables and  $y$  values are dependent variables

In the form of matrix,

$$\begin{matrix} & Y & & X & & \beta \\ \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} & \approx & \begin{bmatrix} 1 + x_{11} + x_{21} + \dots + x_{n1} \\ 1 + x_{12} + \dots + x_{n2} \\ \vdots \\ 1 + x_{1n} + \dots + x_{nn} \end{bmatrix} & \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} & + \epsilon \end{matrix}$$

where  $\beta = ((X^T X)^{-1} X^T) Y$

The above values can be used to plot the best fit line and can be used to predict future values.