Linear Regression Using Gradient Descent

* Batch gradient Descent

Multiple linear Regression
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1 \ldots$

for

cgpa	iq	lap
8.1	9.3	3.2
7.8	95	3.5

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

Step 1 :- Random values
$\beta_0 = 0, \quad \beta_1, \beta_2 = 1$

Step 2 :- epoch (Iteration) = 100,
learning rate = 0.1 (η)

$\beta_0 = \beta_0 - \eta \text{ slope}$

$$l = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= \frac{1}{2} \left[(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \right]$$

$$= \frac{1}{2} [y_i - \beta_0 - \beta_1 x_{11} - \beta_2 x_{12}]^2 +$$

$$[y_2 - \beta_0 - \beta_1 x_{21} - \beta_2 x_{22}]^2$$

$$\frac{\partial l}{\partial \beta_0} = \frac{1}{2} [2(y_i - \hat{y}_1)(-1)] + [2(y_2 - \hat{y}_2)(-1)]$$

Page No.
Date

$$\frac{\partial L}{\partial \beta_0} = -\frac{2}{2}\left[(Y_i - \hat{Y}_1) + (Y_2 - \hat{Y}_2)\right]$$

Consider n rows.

$$\frac{\partial L}{\partial \beta_0} = -\frac{2}{n}\left[(Y_1 - \hat{Y}_1) + (Y_2 - \hat{Y}_2) \dots (Y_n - \hat{Y}_n)\right]$$

$$\boxed{\frac{\partial L}{\partial \beta_0} = -\frac{2}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)}$$ for β_0

for β_1

$$L = \frac{1}{2}\sum (Y_i - \hat{Y}_i)^2$$

$$= \frac{1}{2}\Big[(Y_1 - \beta_0 - \beta_1 X_{11} - \beta_2 X_{12})^2$$

$$+ (Y_2 - \beta_0 - \beta_1 X_{21} - \beta_2 X_{22})^2\Big]$$

$$\frac{\partial L}{\partial \beta_1} = \frac{1}{2}\left[2(Y_1 - \hat{Y}_1)(-X_{11}) + 2(Y_2 - \hat{Y}_2)(-X_{21})\right]$$

$$= -\frac{2}{2}\left[(Y_1 - \hat{Y}_1)(X_{11}) + (Y_2 - \hat{Y}_2)(X_{21})\right]$$

for n rows.

$$= -\frac{2}{n}\Big[(Y_1 - \hat{Y}_1)(X_{11}) + (Y_2 - \hat{Y}_2)(X_{21})$$

$$\dots\Big]$$

$$\boxed{\frac{\partial L}{\partial \beta_1} = -\frac{2}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(X_{i1})}$$

$$\boxed{\frac{\partial L}{\partial \beta_2} = -\frac{2}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(X_{i2})}$$

for all Column and n rows.

$$\boxed{\frac{\partial L}{\partial \beta_m} = \frac{-2}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)\, x_{im}}$$

$\beta_{new} = \beta_{old} - \eta$ Slope

To calculate $\hat{y}_i$

$$\hat{y}_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13}.$$

$$\hat{y}_1 = \beta_0 + [x_{11}\ x_{12}\ x_{13}] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

for n row or all values.

$$\hat{y}_i = \beta_0 + [x_{11}\ x_{12}\ x_{13} \cdots] \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \end{bmatrix}$$

- for Coefficient

$$\frac{\partial L}{\partial \beta_1} = \frac{-2}{n} \cdot \sum (y_i - \hat{y}_i)(x_{i1})$$

Consider

x_1	x_2	y	$\hat{y}$
1	2	5	6
3	4	7	8

$$y_i - \hat{y}_i = [5 \cdot 7] - [6 \cdot 8]$$
$$= [-1 \quad -1]$$

$$[-1 \quad -1] \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$
$$[-1 \quad -3]$$
$$[-4]$$

for β_1 →

$$\boxed{\therefore \frac{\partial L}{\partial \beta_m} = \frac{-2}{n} \Sigma (y_i - \hat{y}_i) x_{im}}$$

* In stochastic Gradient descent

we calculate coefficient for every row and update it as it require Less number of epoches.

$$\frac{\partial L}{\partial \beta_0} = \frac{-2}{n} \Sigma (y_i - \hat{y}_i)$$

for stochastic GD.

$$\boxed{\frac{\partial L}{\partial \beta_0} = -2 (y_i - \hat{y})}$$

(calculate $\hat{y}$ for every random row)

$$\frac{\partial L}{\partial \beta_1} = -\frac{2}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)(x_{i1})$$

for stochastic GD.

$$\boxed{\frac{\partial L}{\partial \beta_1} = -2 (y_i - \hat{y}_i)(x_{i1})}$$

~~Single~~

Mini - Batch Gradient descent

Mini Batch Gradient descent work exactly like Batch gradient descent But instead of loading all the data at one time. In Mini Batch gradient descent we load data in mini Batches.