BATCH GRADIENT DESCENT

Lets assume we have a dataset with n input columns

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots\ldots \beta_n x_n$$

so the total number of variables to calculate are n+1

1. Take random values of $\beta_0, \beta_1, \beta_2 \ldots\ldots \beta_n$

$$\beta_0 = 0 \quad , \beta_1, \beta_2 \ldots\ldots \beta_n = 1$$

2. Epoch = 100, lr = 0.01

$$\beta_0 = \beta_0 - \eta \text{ slope} \rightarrow \text{slope} = \frac{\partial L}{\partial \beta_0}$$

$$\beta_1 = \beta_1 - \eta \text{ slope}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\beta_m = \beta_n - n \text{ slope}$$

where slope is the partial derivative of loss function wrt value of each β_0 to β_m.

we know,

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

$$\Rightarrow \frac{1}{n} \left[(y_1 - \hat{y}_1]^2 + (y_2 - \hat{y}_2)^2 \ldots (y_n - \hat{y}_n)^2 \right]$$

$$\Rightarrow \frac{1}{n} \left[(y_1 - \beta_0 - \beta_1 X_{11} \ldots\ldots \beta_n X_{nm})^2 + (y_2 - \beta_0 - \beta_1 X_{21} \ldots \ldots X_{2m}) \right]$$

$$\frac{\partial L}{\partial \beta_0} = \frac{1}{n}\left[2(y_1 - \hat{y}_1)(-1) + 2(y_2 - \hat{y}_2)(-1) + 2(y_n - \hat{y}_n)(-1)\right]$$

DERIVATIVE

$$\Rightarrow \frac{-2}{n}\left[(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + \ldots\ldots (y_n - \hat{y}_n)\right]$$

$$\boxed{\Rightarrow \frac{-2}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) = \frac{\partial L}{\partial \beta_0}}$$

Now again for β_1 slope

$$L = \frac{1}{n}\left[(y_1 - \beta_0 - \beta_1 x_{11} \ldots\ldots \beta_n x_{1m}) + (y_n - (\beta_0 + \beta_1 x_{nm} \ldots\ldots \beta_n x_{nm})\right]$$

DERIVATIVE

$$\frac{\partial L}{\partial \beta_1} = \frac{1}{n}\left[2(y_1 - \hat{y}_1)(-x_{11}) + 2(y_2 - \hat{y}_2)(-x_{21}) \ldots \ldots\ldots 2(y_n - \hat{y}_n)(-x_{n1})\right]$$

$$\frac{\partial L}{\partial \beta_1} = \frac{2}{n}\left[\cancel{2}(y_1 - \hat{y}_1)(-x_{11}) \ldots\ldots (y_n - \hat{y}_n)(-x_{n1})\right]$$

$$\boxed{\Rightarrow \frac{-2}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)\, x_{i2} = \frac{\partial L}{\partial \beta_1}}$$

Similarly,

$$\boxed{\begin{aligned} \frac{\partial L}{\partial \beta_2} &= \frac{-2}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)\, x_{i2} \\ \frac{\partial L}{\partial \beta_m} &= \frac{-2}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)\, x_{im} \end{aligned}}$$