Jeet Banerjee

Mathematical Formulation
(Batch Gradient Descent + Multiple Linear Regression)

In batch gradient descent for MLR, we need to iteratively update β_0 (offset) and $\beta_1, \beta_2, \ldots, \beta_m$ (coefficients) as per the following equations:

$$\boxed{\begin{aligned} \beta_0 &= \beta_0 - \eta * slope(L)_{\beta_0} \\ \beta_1 &= \beta_1 - \eta * slope(L)_{\beta_1} \\ \beta_2 &= \beta_2 - \eta * slope(L)_{\beta_2} \\ &\vdots \\ \beta_m &= \beta_m - \eta * slope(L)_{\beta_m} \end{aligned}} \quad \text{—(3)}$$

where η = learning rate

L = loss function

$slope(L)_{\beta_i}$ = slope of Loss function w.r.t β_i

Let's calculate the loss function (L) and respective slopes for eq. (3):

$$\boxed{L = \text{loss function} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad \text{—(4)}$$

$\Rightarrow$ denotes MSE (mean squared error)

From eq. (1), $\hat{y}_i$ can be written as:

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_m x_{im}$$

Hence,

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3} - \cdots - \beta_m x_{im})^2$$

Now,

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

$$\Rightarrow \boxed{\frac{\partial L}{\partial \beta_0} = -\frac{2}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)} \quad \text{—(5a)}$$

Similarly,

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

$$\Rightarrow \boxed{\frac{\partial L}{\partial \beta_1} = -\frac{2}{n}\sum_{i=1}^{N}(y_i - \hat{y}_i)x_{i1}} \quad \text{—(5b)}$$

Next, $\frac{\partial L}{\partial \beta_2} = \frac{1}{n}\sum_{i=1}^{N} 2(y_i - \hat{y}_i)(-x_{i2})$

$$\Rightarrow \boxed{\frac{\partial L}{\partial \beta_2} = -\frac{2}{n}\sum_{i=1}^{N}(y_i - \hat{y}_i)x_{i2}} \quad \text{—(5c)}$$

Jeet Banerjee

Mathematical Formulation
(Batch Gradient Descent + Multiple Linear Regression)

Consider, in the given dataset,

$n \rightarrow$ no. of rows / no. of data points

$m \rightarrow$ no. of i/p or predictor variables

and there is single o/p

Hence, the dataset can be look like:

x_1	x_2	x_3	...	x_m	y
x_{11}	x_{12}	x_{13}	...	x_{1m}	y_1
x_{21}	x_{22}	x_{23}	...	x_{2m}	y_2
⋮	⋮	⋮	⋮	⋮	⋮
x_{n1}	x_{n2}	x_{n3}	...	x_{nm}	y_n

Now, for MLR, the predicted o/p can be written as:

$$\boxed{\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_m x_m} \quad -(1)$$

— where β_0 = intercept / offset

$\beta_1, \beta_2, \beta_3, \ldots, \beta_m$ = coefficients, linked to 'm' no. of i/p variables

Elaborating eq.(1), we get:

$$\hat{y}_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_m x_{1m}$$

$$\hat{y}_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_m x_{2m}$$

$$\hat{y}_3 = \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} + \beta_3 x_{33} + \cdots + \beta_m x_{3m}$$

$$\vdots$$

$$\hat{y}_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \cdots + \beta_m x_{nm}$$

In concise,

$$\boxed{\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \beta_0 + \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1m} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2m} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3m} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nm} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_m \end{bmatrix}} \quad -(2)$$

Jeet Banerjee

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

$$\Rightarrow \boxed{\frac{\partial L}{\partial \beta_m} = -\frac{2}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)(x_{im})} \quad -(5d)$$

Hence, summarizing equations in (5),

$$\text{slope}(L)_{\beta_0} = -\frac{2}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)$$

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

→ denotes first i/p column

$$\text{slope}(L)_{\beta_2} = -\frac{2}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)x_{i2} \quad -(6)$$

→ denotes second i/p column

$$\vdots$$

$$\text{slope}(L)_{\beta_m} = -\frac{2}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)x_{im}$$

→ denotes mth i/p column

→ Where $\hat{y}$ is denoted in eq. (2)

Implementation steps

- Calculate $\hat{y}$ using eq. (2)
- calculate slopes of loss function using eq. (6)
- update the intercept and coefficients using eq. (3)