Formative 3 – Probability Distributions, Bayesian Probability, and Gradient Descent Implementation

$$y = mx + b$$

Using data points $(1,3)$ and $(3,6)$, with:

- Initial $m = -1$
- Initial $b = 1$
- Learning rate $\alpha = 0.1$

Cost function is Mean Squared Error (MSE)

Predicted Value:

$$\hat{y}_i = m x_i + b$$

Gradients (for MSE):

$$\frac{\partial J}{\partial m} = -\frac{2}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i) x_i$$

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

GD update:

$$m_{new} = m_{old} - \alpha \frac{\partial J}{\partial m}$$

$$b_{new} = b_{old} - \alpha \frac{\partial J}{\partial b}$$

First Iteration

$$m = -1, \quad b = 1$$

for $(1,3)$: $\hat{y}_1 = (-1) \times 1 + 1 = 0$

for $(3,6)$: $\hat{y}_2 = (-1) \times 3 + 1 = -2$

Errors: $3 - 0 = 3$

$6 - (-2) = 8$

Gradients: $\frac{\partial J}{\partial m} = -1 \times (3 \times 1 + 8 \times 3) = -1 \times (3 + 24) = -27$

$$\frac{\partial J}{\partial b} = -1 \times (3 + 8) = -11$$

Update:

$$m_{new} = -1 + 2.7 = 1.7$$

~~m old =~~

$$b_{new} = 1 + 1.1 = 2.1$$

Second iteration

$m = 1.7,\ b = 2.1$

For $(1,3)$: $\hat{y}_1 = 1.7\times 1 + 2.1 = 3.8$

For $(3,6)$: $\hat{y}_2 = 1.7\times 3 + 2.1 = 7.2$.

For errors:

- $3 - 3.8 = -0.8$
- $6 - 7.2 = -1.2$

Gradients:

$$\frac{\partial J}{\partial m} = -1\times(-0.8\times 1 + -1.2\times 3)$$
$$= -1\times(-0.8 - 3.6) = 4.4$$

$$\frac{\partial J}{\partial b} = -1\times(-0.8 - 1.2) = 2.0.$$

Update

$$m_{new} = 1.7 - 0.44 = 1.26$$
$$b_{new} = 2.1 - 0.2 = 1.9$$

Third ~~up~~ Iteration.

$m = 1.26 \qquad b = 1.9$

For $(1,3)$: $\hat{y}_1 = 1.26 \times 1 + 1.9 = 3.16$

For $(3,6)$: $\hat{y}_2 = 1.26 \times 3 + 1.9 = 5.68$

Errors

- $3 - 3.16 = -0.16$
- $6 - 5.68 = 0.32$

$$\frac{\partial J}{\partial m} = -1 \times (0.16 \times 1 + 0.32 \times 3).$$
$$= -1 \times (-0.16 + 0.96) = -0.8$$

$$\frac{\partial J}{\partial b} = -1 \times (-0.16 + 0.32) = -0.16$$

Update : $m_{new} = 1.26 + 0.08 = 1.34$

$b_{new} = 1.9 + 0.016 = 1.916$

Fourth iteration

for (1,3): $\hat{y}_1 = 1.34 \times 1 + 1.916 = 3.256$

for (3,6): $\hat{y}_2 = 1.34 \times 3 + 1.916 = 4.02 + 1.916 = 6.036$

For errors:

* $3 - 3.256 = -0.256$
* $6 - 6.036 = -0.036$

Now, gradients:

$$\frac{\partial J}{\partial m} = -1 \times [(-0.256) \times 1 + (-0.036) \times 3] = -1 \times (-0.256 - 0.108)$$
$$= 0.364$$

$$\frac{\partial J}{\partial b} = -1 \times (-0.256 - 0.036) = 0.292$$

Update

$m_{new} = 1.34 - 0.1 \times 0.364 = 1.34 - 0.0364 = 1.3036$

$b_{new} = 1.916 - 0.1 \times 0.292 = 1.916 - 0.0292 = 1.8868$

Trend & Conclusion.

* With each step, values of m and b change to reduce prediction error

* The update size also shrinks as we approach a better fit for the data (converging).