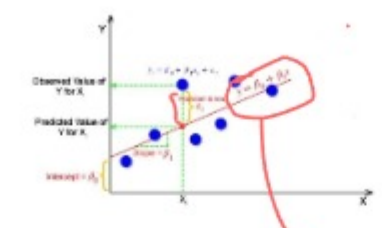


- points = b
- y = Dependent variable (output)
 - x = Independent variable (input)
 - m = Slope of the line
 - b = Intercept



this is formula for error

$\epsilon_i = y_{\text{predicted}} - y_i$

Cost function
it tells us the optimal values for the coefficients which provides the best fit line

Mean squared error

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

Handwritten notes: y_i = Observed value, \hat{y}_i = Predicted value, n = Total number of features.

Gradient descent
It is an optimization algo that optimize the cost function to reach the minimum value



Learning Rate

$b_0 = b_0 - \alpha \cdot \frac{2}{n} \sum_{i=1}^n (pred_i - y_i)$ This is for intercept value

$b_1 = b_1 - \alpha \cdot \frac{2}{n} \sum_{i=1}^n (pred_i - y_i) \cdot x_i$ This is for your feature x_1

Handwritten notes: α = Learning Rate, b_0 = old value, b_1 = new value.

Observed x_i	Observed y_i
1	20
2	40
3	60
4	80

Step 1
 $p(y) = X_1b_1 + b_0$
 $b_1, b_2 = 0$
learning rate = 0.01
 $p(y) = 0$

Step 2
 $MSE = \frac{1}{4} [(0-20)^2 + (0-40)^2 + (0-60)^2 + (0-80)^2]$
 $MSE = 3000$

Step 3
Gradient descent

Handwritten notes: $\alpha = 0.01$, $\sum (y_i - \hat{y}_i) = -600$, $\sum (y_i - \hat{y}_i) \cdot x_i = -300$.

	x_i	y_i	\hat{y}_i	$(y_i - \hat{y}_i)$	$(y_i - \hat{y}_i) \cdot x_i$
1	20	20	0	20	400
2	40	40	0	40	1600
3	60	60	0	60	3600
4	80	80	0	80	6400

Handwritten calculations:

$$\sum (y_i - \hat{y}_i) = -200 - 400 - 600 - 800 = -2000$$

$$\sum (y_i - \hat{y}_i) \cdot x_i = -100 - 400 - 1800 - 6400 = -8700$$

$$b_1' = 0.01 \cdot (-8700) = -0.87$$

$$b_0' = 0.01 \cdot (-2000) = -0.02$$

$$y' = 0.87x + 0.02$$

Step 4
Repeat from Step 2 for y'