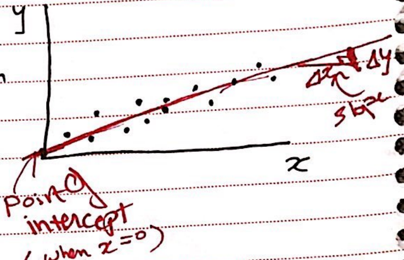
- we try to teach Computer to find the best line that pass through these Points

- a line is represented by slope and point of intercept (intersection with the y axis)



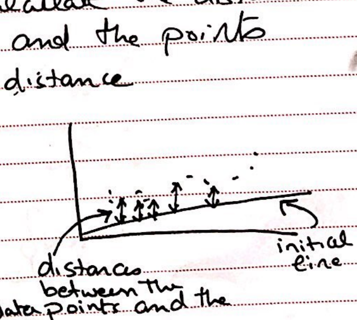
equation of straight line

$$y = a x + b$$

- We start by suggesting any random values for a and b
  the main idea is that we keep enhancing the values of a and b iteratively until reach the best values (the values that make the line the closest to the given data points) (optimization)

- therefore each iteration we calculate the distance between the suggested line and the points and try to minimize that distance

- to get the distance between the points and the line suggested we substitute in the equation of the suggested line



$$\text{Prediction} = a_1 x + b_1$$

Such that a_1 and b_1 are the initial random suggested values for the line

NO.

DATE.

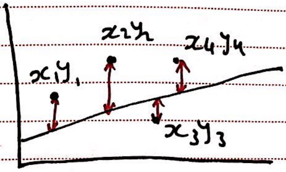
- We are calculating the error between the points on the suggested line and the date points (تعويض مباشر في معادلة الخط)

mean square error

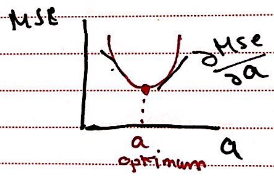
$$MSE = \frac{1}{2n}\Big(y_1 - \underbrace{(ax_1 + b)}_{\text{predicted } y \text{ (point on the line)}}\Big)^2 + \big(y_2 - (ax_2 + b)\big)^2 + \dots + \big(y_n - (ax_n + b)\big)^2$$

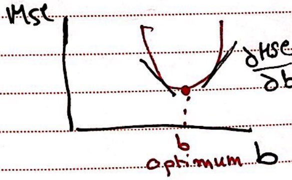
n: number of data points

- We want to minimize the red lines (distances between each data point and its corresponding point on the line)



- In other words, we want to find the minimum value of the MSE i.e find the best values of a and b that give us the min value of MSE





So we consider MSE to be function of a and function of b and through partial derivatives we find the best (optimum) value of a and of b

3

NO.

DATE.

$$\frac{\partial Mse}{\partial a} = \frac{1}{2n}\Big[2.(y_1 - (ax_1 + b))(-x_1) + 2(y_2 - (ax_2 + b))(-x_2) + \ldots \ldots + 2(y_n - (ax_n + b))(-x_n)\Big]$$

Solving the above equation

$$\therefore \frac{\partial Mse}{\partial a} = -\frac{1}{n}(\text{all } y_{data} - \text{all } y_{prediction})(\text{all } x_{data})$$

$$= \frac{1}{n}(\text{all } y_{prediction} - \text{all } y_{data})(\text{all } x_{data})$$

Similarly $\frac{\partial Mse}{\partial b}$ is calculated in the same manner and equal =

$$(\text{all } y_{prediction} - \text{all } y_{data})$$

$$a_{new} = a_{old} - Lr\left(\frac{\partial Mse}{\partial a}\right)_{at\ a\ old}$$

$$b_{new} = b_{old} - Lr\left(\frac{\partial Mse}{\partial b}\right)_{at\ b\ old}$$