



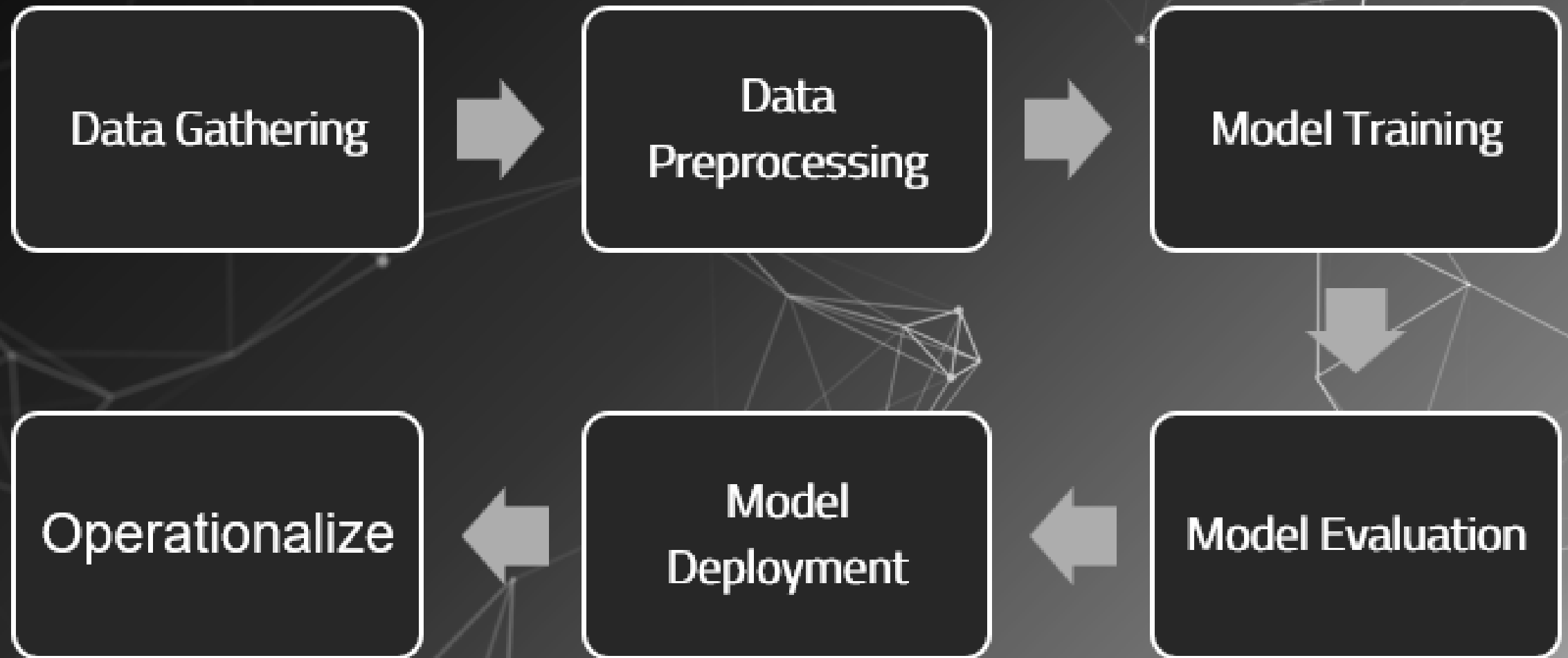
SUPERVISED MACHINE LEARNING

ENG- MOHAMED KHALED IDRIS

ENG- MAYAR SWILAM



Machine Learning Process

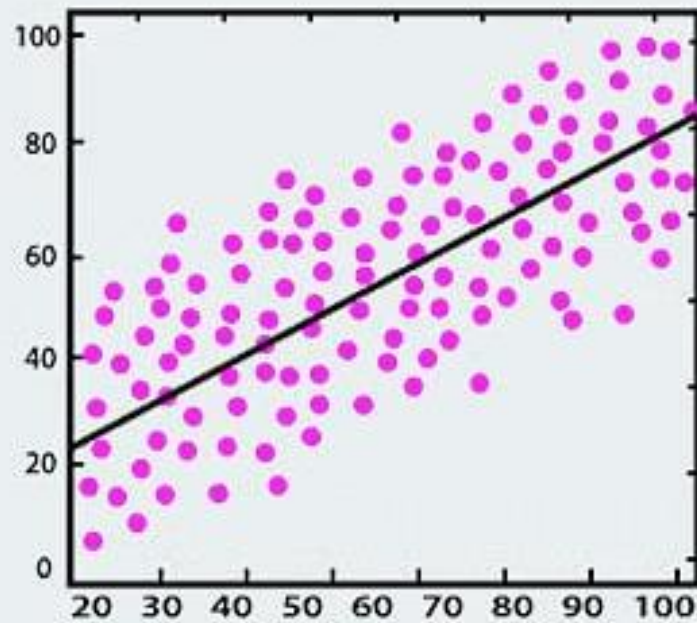


SUPERVISED MACHINE LEARNING (REGRESSION)

WHAT IS REGRESSION ?

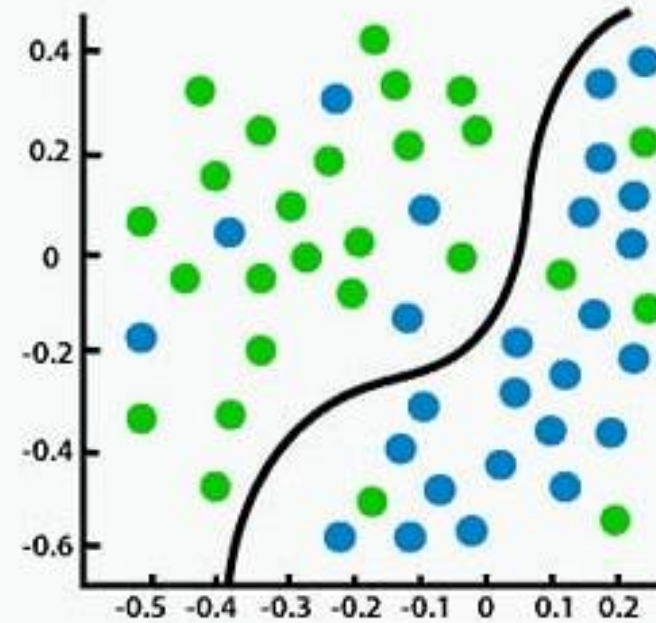
- **Regression** is a statistical technique used in data analysis to model the relationship between a **dependent variable** (also called the outcome or target variable) and **one or more independent variables** (also known as predictor variables or features).
- **The main goal** of regression analysis is to understand and quantify how changes in the independent variables are associated with changes in the dependent variable.

In simpler terms, regression helps us predict or estimate a continuous numerical outcome based on one or more input variables.



Regression

versus



Classification

EXAMPLE

x	1	2	3	4	5	6
y	10	15	20	25	??	??

SUGGEST A WAY TO SOLVE
FOR $X = 5$, $X = 6$

ONE OF THE IDEAS

$$y = mx + b$$

1

Try to fit the previous points on a line

2

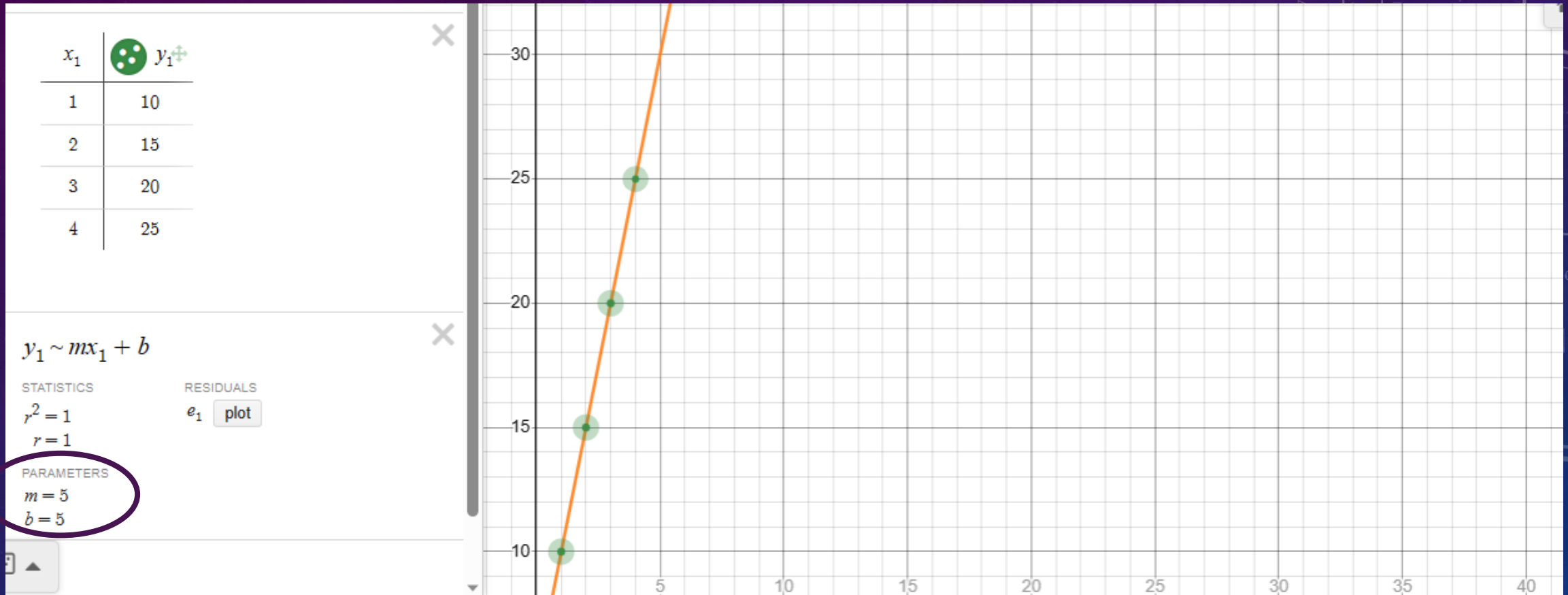
Use the slope equation and find the slope
 $m = y_2 - y_1 / x_2 - x_1$

3

Find the y-intercept e.g. b of the equation

4

Substitute x values on the equation to find the y values



$$Y = 5X + 5$$

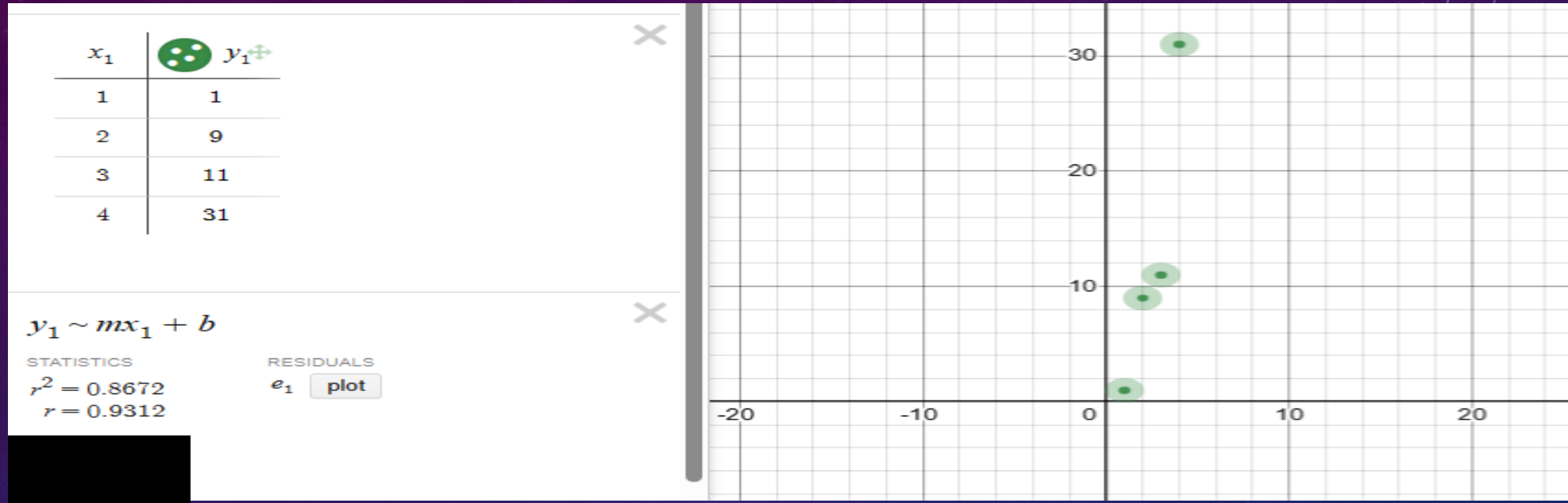
x	1	2	3	4	5	6
y	10	15	20	25	30	??

$$Y = 5X + 5$$

x	1	2	3	4	5	6
y	10	15	20	25	30	35

WHAT IF

x	1	2	3	4	5	6
y	1	9	11	31	??	??



THE ENTIRE PROBLEM

THE IDEA

- In linear regression, the goal is to **FIND THE BEST-FITTING STRAIGHT LINE (LINEAR EQUATION)** that represents the relationship between the predictor(s) and the target variable.

$$y = mx + b$$

Where:

- y is the dependent variable (target).
- x is the independent variable (predictor).
- m is the slope of the line, representing how much y changes for a unit change in x .
- b is the y -intercept, indicating the value of y when x is 0.

WHAT IS REGRESSION USED FOR ?



Prediction: Given new input values, regression models can be used to predict the likely value of the dependent variable.



Inference: Regression analysis can help us understand the relationship between variables and provide insights into how changes in the predictors affect the outcome.

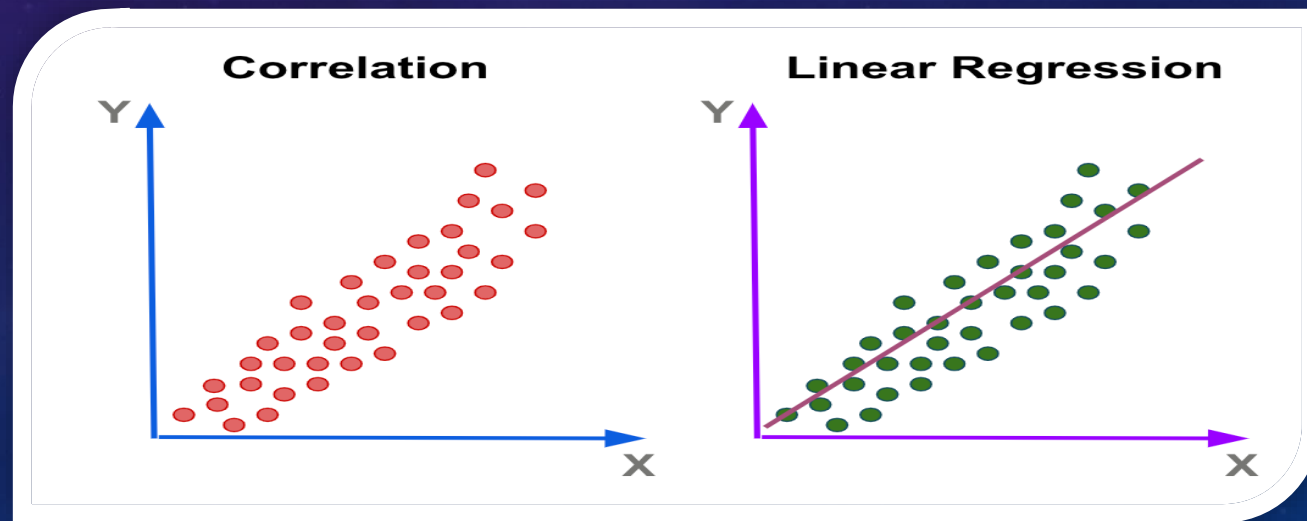


Control: In experimental settings, regression can be used to determine how changes in the independent variables impact the dependent variable, allowing for better control over the experimental conditions.

LINEAR REGRESSION

LINEAR REGRESSION

- **Linear regression** is a statistical method used to model the relationship between a dependent variable (target) and one or more independent variables (predictors) by assuming a linear relationship between them.
- It's one of the simplest and most commonly used regression techniques.



THETA0= 5 , THETA 1 = 2 , H(X) = 5 + 2X

J = 1/ 14 (0+1+4+4+4+25+9)= 47/14=3.3

$$J(\theta) = h(x) - y$$

$$J(\theta) = (h(x) - y)^2$$

$$J(\theta) = \sum_{i=1}^m (h(x^i) - y^i)^2$$

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m (h(x^i) - y^i)^2$$

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h(x^i) - y^i)^2$$

x	y	h(x)	h(x) - y	(h(x) - y)^2
1	7	7	0	0
2	8	9	1	1
2	7	9	2	4
3	9	11	2	4
4	11	13	2	4
5	10	15	5	25
5	12	15	3	9

METHODS TO FIND THE BEST FITTING-LINE

Ordinary Least Squares (OLS)

Gradient Descent (GD)

Ridge

Lasso

Elastic Net

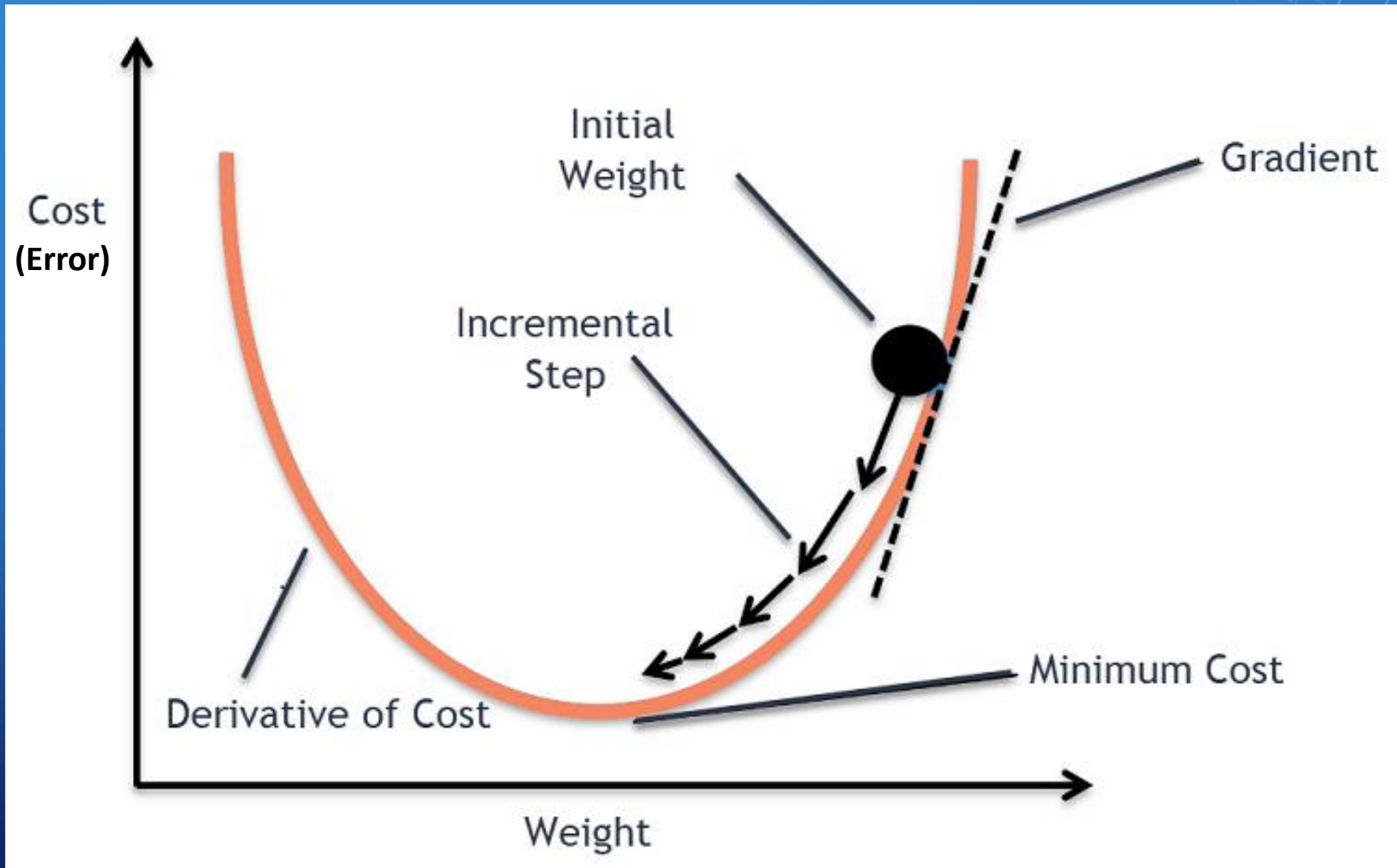
Stochastic Gradient Decent (SGD)

GRADIENT DESCENT

- **Gradient descent** is an iterative optimization algorithm used to find the minimum of a function, particularly in machine learning and numerical optimization.
- It's widely used for updating the parameters of a model in order to minimize a ***COST FUNCTION*** or objective function.
- The "gradient" in gradient descent refers to the gradient vector of **PARTIAL DERIVATIVES** of the function with respect to its parameters.

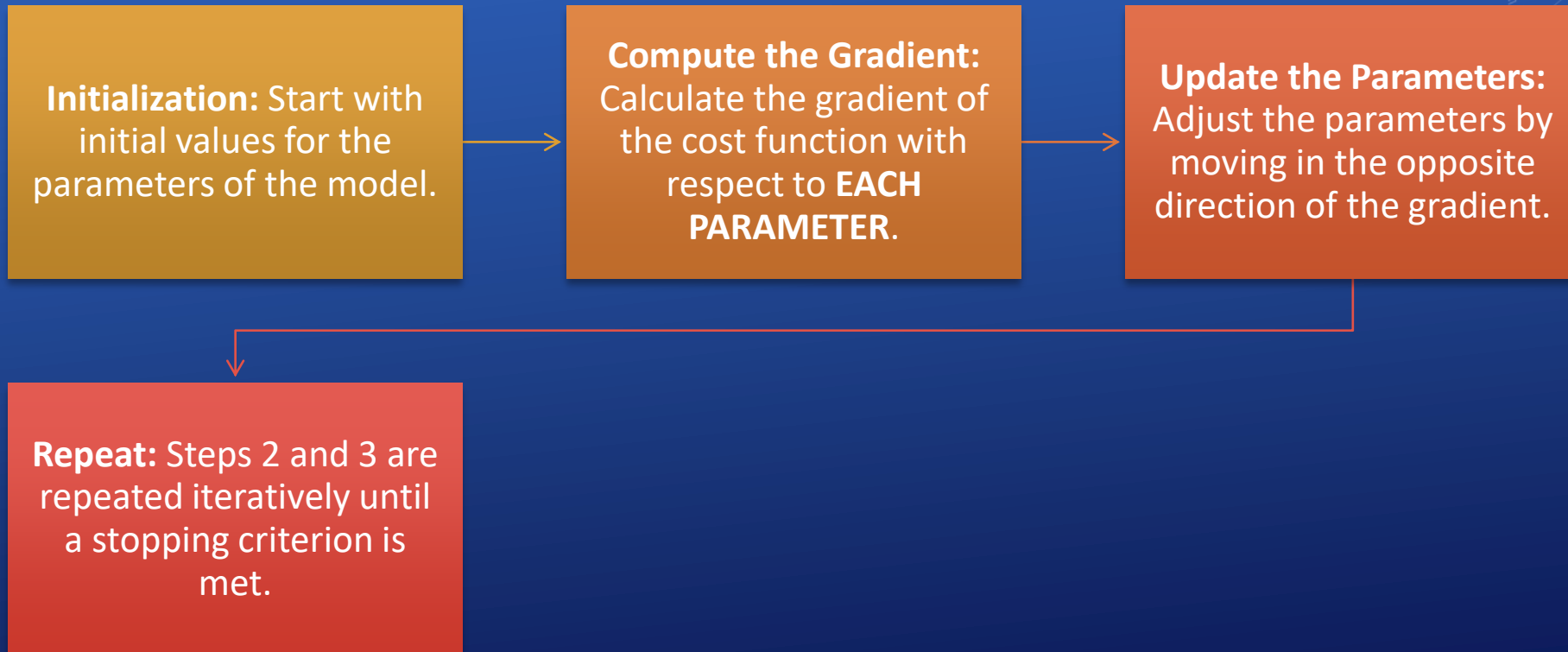
THE BASIC IDEA OF GRADIENT DESCENT

- The fundamental idea behind gradient descent is to **ITERATIVELY** adjust the parameters of a model in the direction of the steepest descent (negative gradient) of the cost function.
- By repeatedly taking steps in the direction that reduces the value of the cost function, the algorithm eventually converges to a local or global minimum.



STEPS OF THE GRADIENT DESCENT

$$\text{New Parameter} = \text{Old Parameter} - \text{Learning Rate} \times \text{Gradient}$$



1. **Batch Gradient Descent (BGD):** In batch gradient descent, the entire dataset is used to compute the gradient of the cost function in each iteration. It provides a precise estimate of the gradient but can be computationally expensive for large datasets.
2. **Stochastic Gradient Descent (SGD):** In stochastic gradient descent, a single randomly chosen data point is used to compute the gradient in each iteration. This approach is faster but can result in noisy updates and slower convergence due to the randomness.
3. **Mini-Batch Gradient Descent:** Mini-batch gradient descent strikes a balance between BGD and SGD by using a small batch of data points for each iteration. It provides a good trade-off between computational efficiency and noise in the gradient estimate.

TYPES OF GRADIENT DESCENT

PUTTING ALL TOGETHER

- We are using (Least Square) and must be added to **COST FUNCTION**
- Our required parameters is M slope and b y-intercept
- We will use **Batch Gradient Descent** in order to modify the M and b parameters as to minimize the cost function (Error rate)
- This process will be iterative until the stopping criteria is met
- The line generated by using best m and b values are **BEST FITTING-LINE**

COST FUNCTIONS

- In linear regression, **the cost function** (also known as the loss function or objective function) quantifies how well the model's predictions match the actual observed values.
- The goal of linear regression is to find the parameters (coefficients) of the linear equation **THAT MINIMIZE THIS COST FUNCTION**.

1. **Mean Squared Error (MSE):** MSE is one of the most widely used cost functions in linear regression. It calculates the average squared difference between the predicted values and the actual values for all data points.

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


Where:

- n is the number of data points.
- y_i is the actual value of the dependent variable for the i th data point.
- \hat{y}_i is the predicted value of the dependent variable for the i th data point.

Minimizing MSE results in the coefficients that provide the best-fitting line in terms of minimizing the sum of squared residuals.

This one will be used as cost function
MEAN SQUARED ERROR (MSE)

2. **Root Mean Squared Error (RMSE):** RMSE is the square root of the MSE and provides a measure of the average error between the predicted and actual values. It is commonly used when you want the error metric to be in the same units as the dependent variable.

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

ROOT MEAN SQUARED ERROR (RMSE)

3. **Mean Absolute Error (MAE):** MAE computes the average absolute difference between the predicted and actual values. It is less sensitive to outliers compared to MSE.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

29

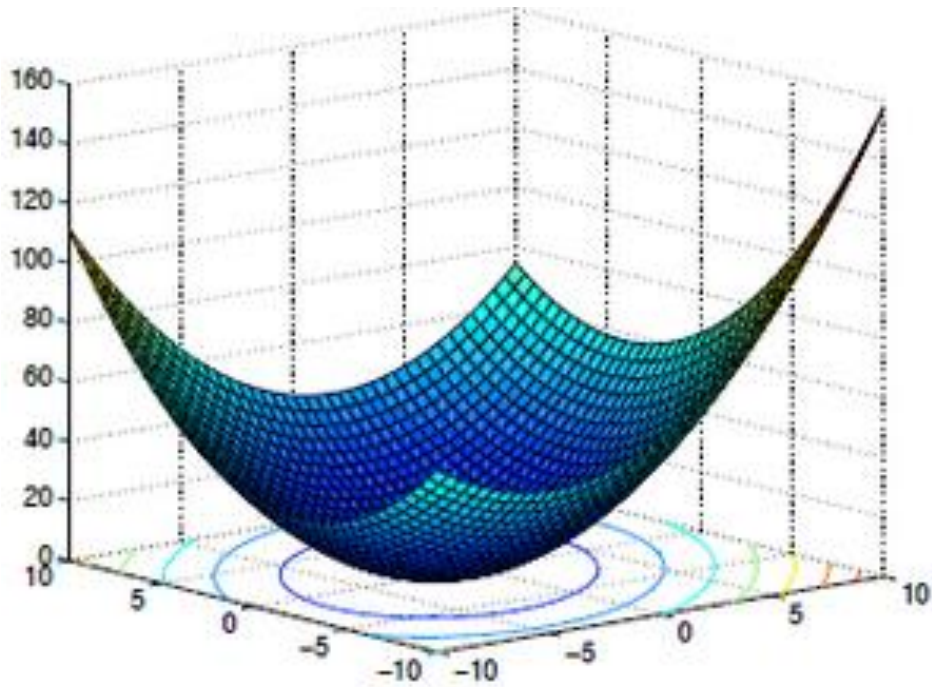
MEAN ABSOLUTE ERROR (MAE)

THE FULL EQUATION

$$J(b_0, b_1) = \frac{1}{2n} \sum_{i=1}^n (y_i - h_i)^2$$

- Where $h = b_0 + b_1x$
- And J is cost function of MSE

THE RELATION BETWEEN J , B_0 AND B_1



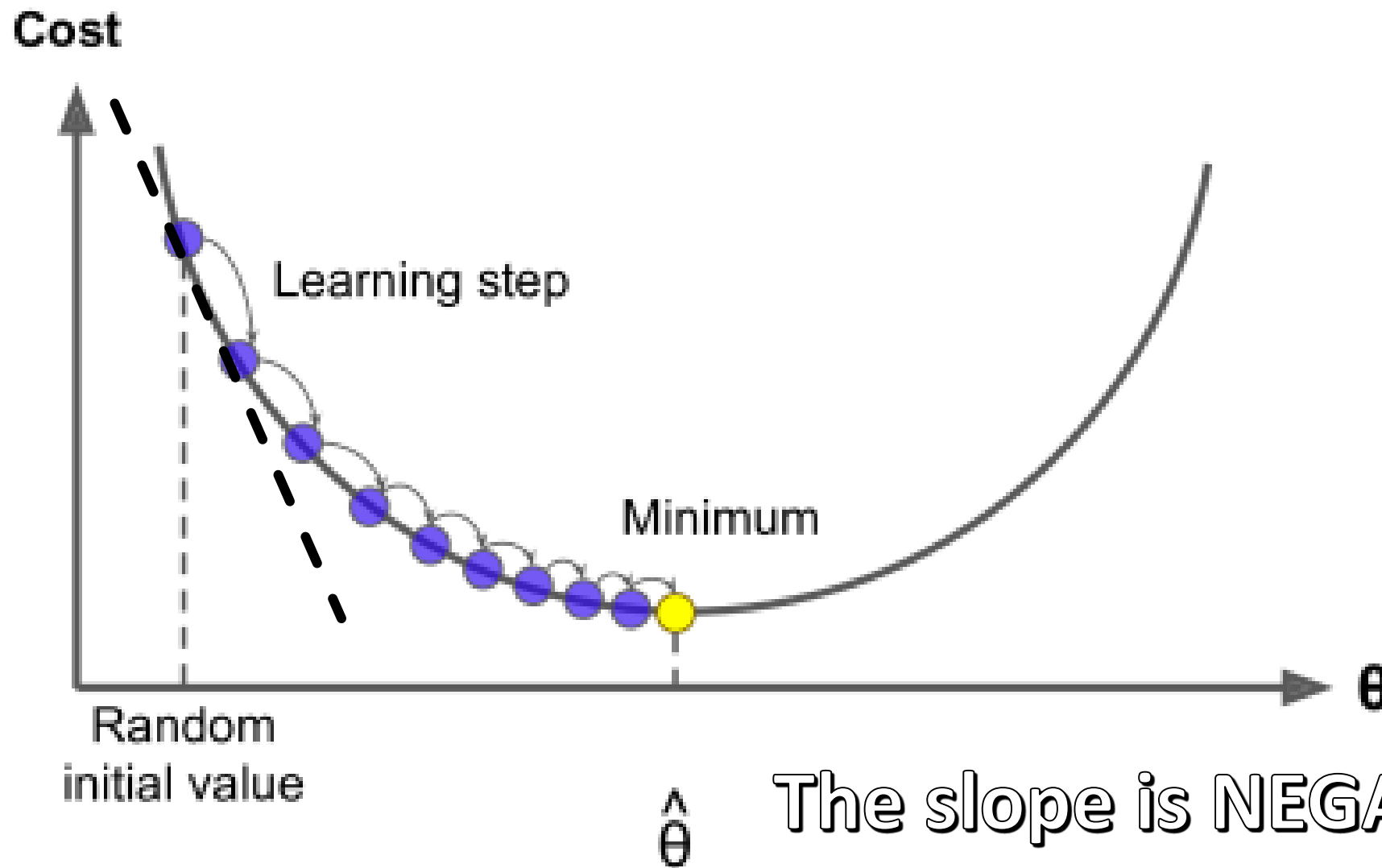
3D

STEPS OF THE GRADIENT DESCENT

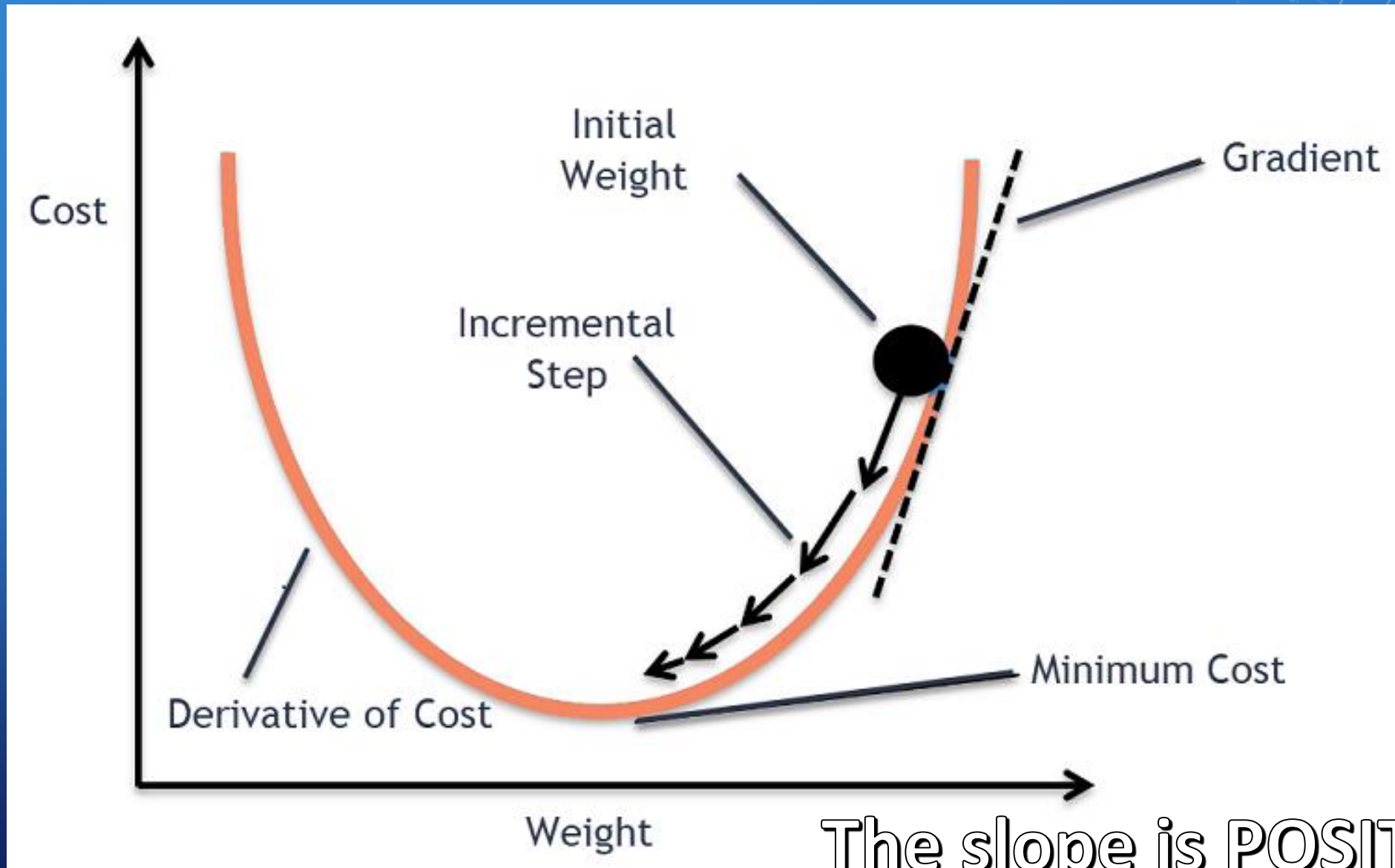
$$\text{New Parameter} = \text{Old Parameter} - \text{Learning Rate} \times \text{Gradient}$$



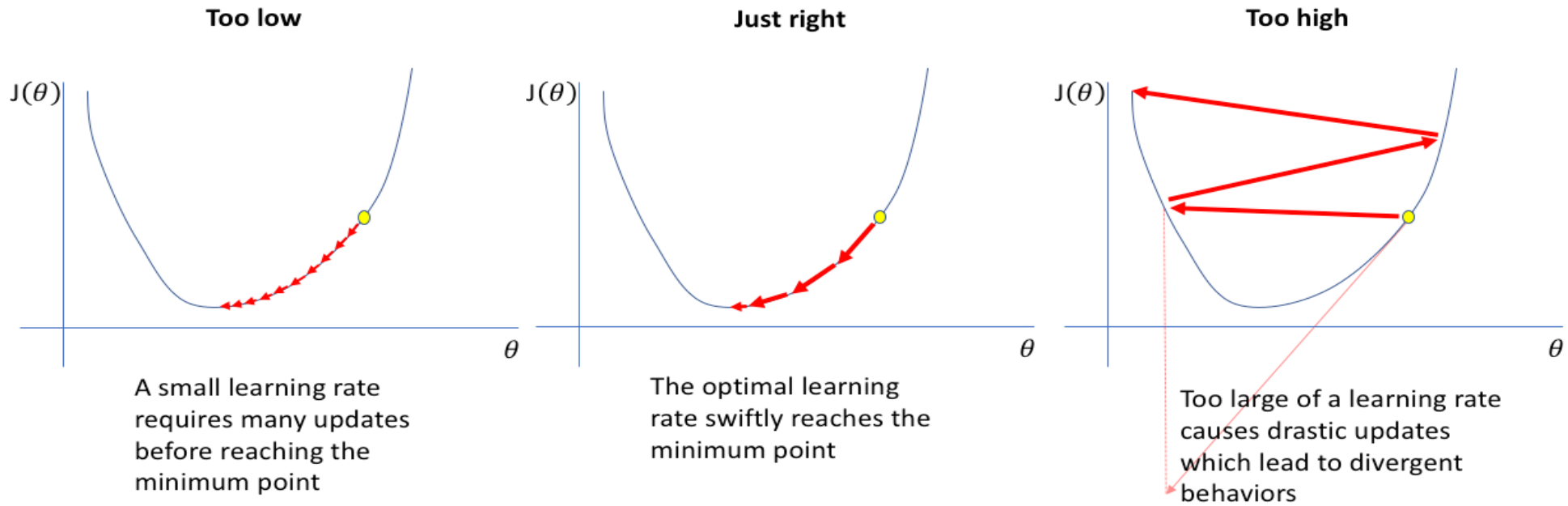
TO REMEMBER



The slope is **NEGATIVE**



The slope is POSITIVE



LEARNING RATE (CONT.)

Area x1	Rooms x2	Age x3	Sea x4	Price
120	3	4	1	150
60	1	15	0	50
100	4	1	1	230

Theta

$$y = \Theta_1 X_1 + \Theta_2 X_2 + \Theta_3 X_3 + \Theta_4 X_4$$

Theta

$$y = \Theta_1 X_1 + \Theta_2 X_2 + \Theta_3 X_3 + \Theta_4 X_4$$

$$Y = 53.6 X_1 + 8547 X_2 + 653.3 X_3 + 452 X_4$$

1. Building Math Equation

2. Randomised Thetas: between 0 and 1

a. Calculating Predicted Value

Phase 1:

$$y = 0.56 X_1 + 0.496 X_2 + 0.87450 X_3 + 0.02036 X_4$$

$$y = 0.56 \cdot 120 + 0.4963 \cdot 3 - 0.874504 \cdot 4 + 0.020361 \cdot 1 = 170$$



QUESTIONS

THANK YOU