AMIT

Machine Learning

Session 1

Agenda:

1	Introduction to Machine Learning
2	Linear Regression
3	Polynomial Regression

Introduction to Machine Learning

What is Machine Learning?

Machine Learning is a branch of artificial intelligence (AI) that focuses on building systems that can learn from and make decisions based on data. Instead of being explicitly programmed to perform a task, ML models identify patterns and learn from the data they are provided.

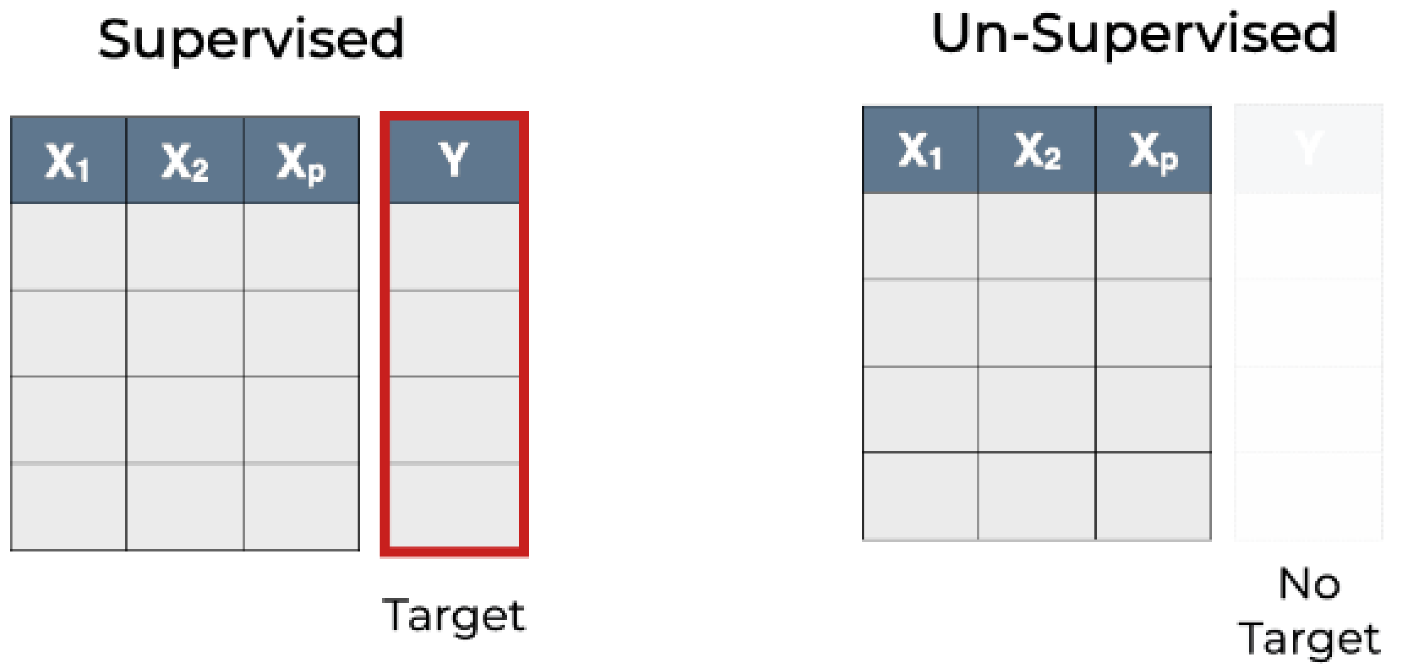
Types of Machine Learning

- **Supervised Learning**: Models are trained on labeled data (i.e., input-output pairs), making predictions based on this training. Examples include classification (e.g., spam detection) and regression (e.g., predicting house prices).

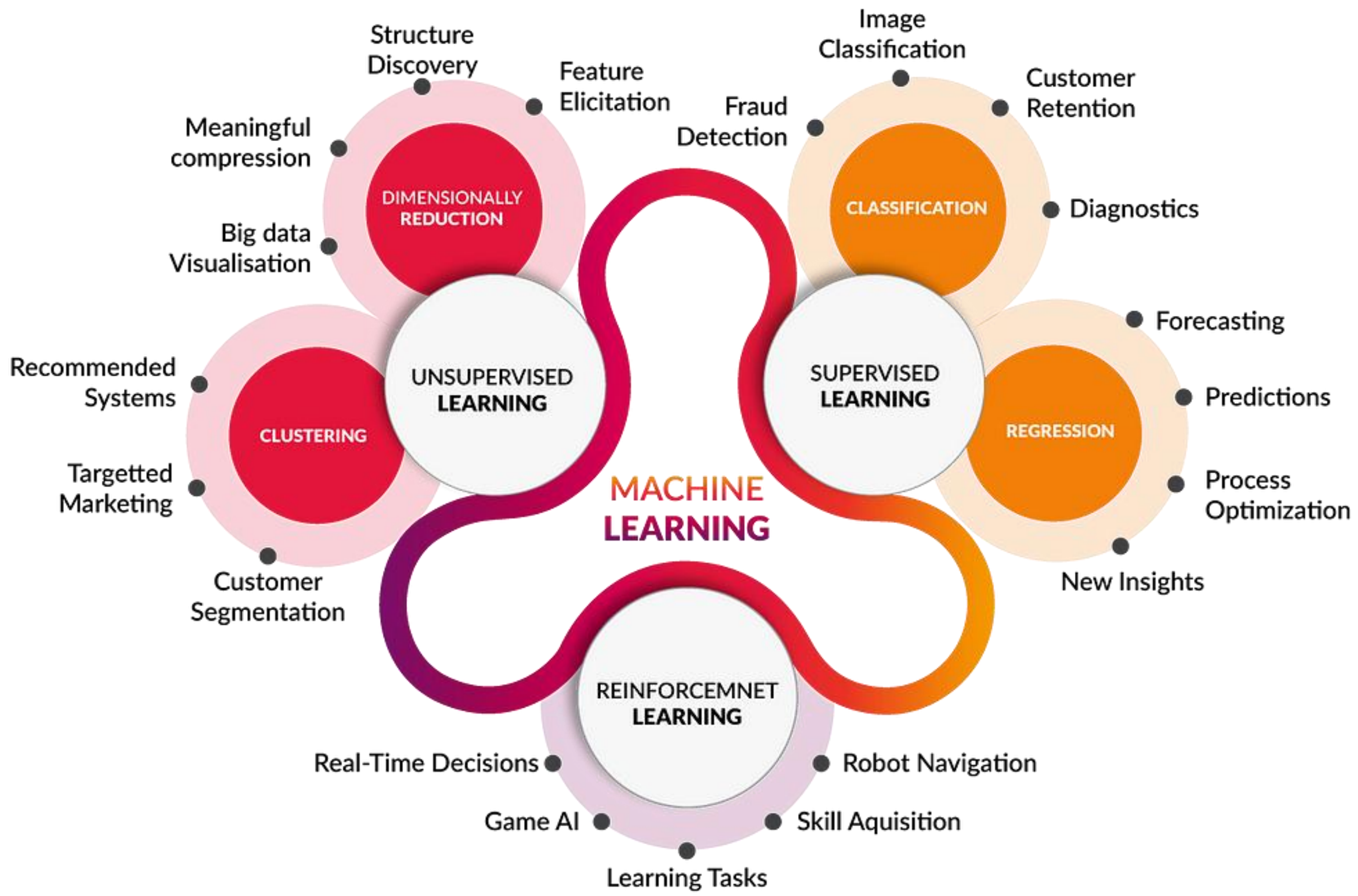
- **Unsupervised Learning**: Models learn from unlabeled data by finding hidden patterns or intrinsic structures in input data. Examples include clustering (e.g., customer segmentation) and association (e.g., market basket analysis).

- **Reinforcement Learning**: Models learn by interacting with an environment and receiving rewards or penalties, like teaching a robot to navigate a maze.

Supervised vs. Unsupervised



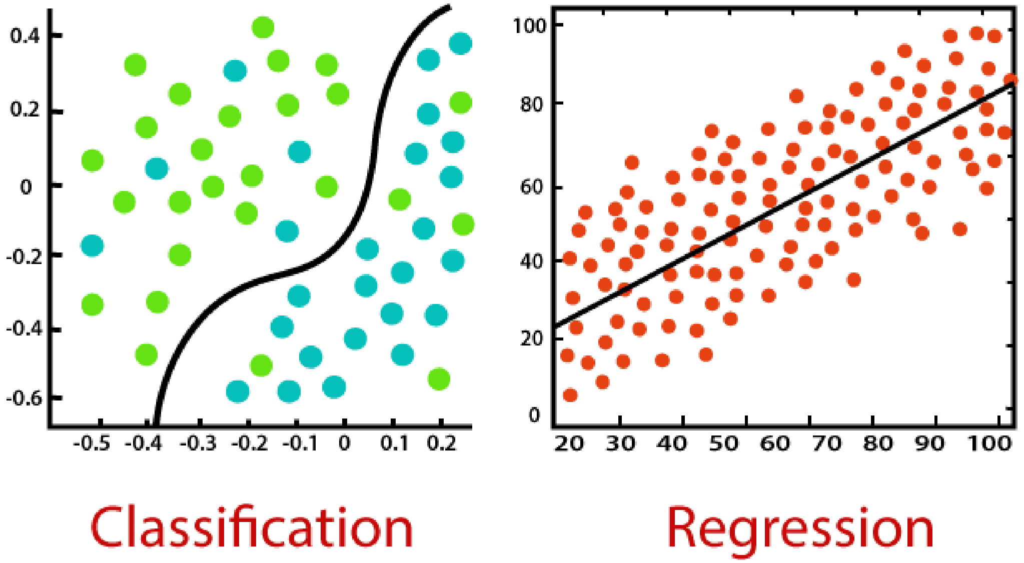
Types of Machine Learning



Key Differences

- **Classification**: The goal is to divide data into different categories using a boundary (e.g., identifying spam emails from non-spam emails).

- **Regression**: The goal is to predict continuous numerical values by fitting a line or curve to the data (e.g., predicting house prices).



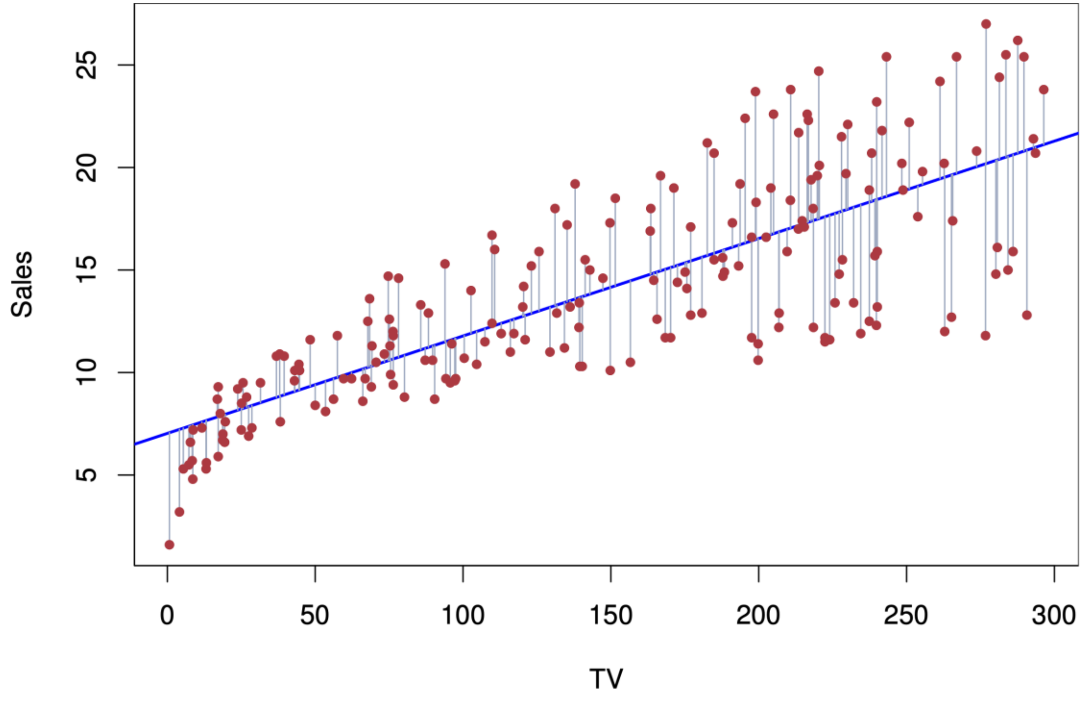
Linear Regression

Linear Regression

Linear regression is one of the simplest and most commonly used predictive modeling techniques. It models the relationship between a dependent (target) variable and one or more independent (predictor) variables by fitting a linear equation to observed data.

Linear Regression

The objective of linear regression is to identify the line of best fit, known as the regression line, that reduces the difference between the predicted values and the actual data points as much as possible.



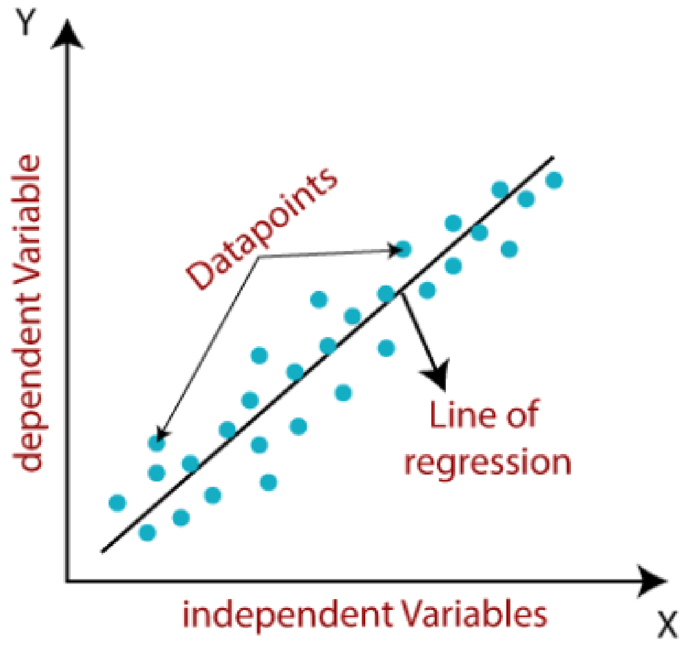
Simple Linear Regression

Simple linear regression uses one independent variable to predict a dependent variable. The relationship is modeled by a straight line

$$y = \beta_0 + \beta_1 x + \epsilon$$

where:

- y is the predicted value (dependent variable).
- x is the independent variable.
- β_0 is the y-intercept (constant term).
- β_1 is the slope (coefficient for x).
- ϵ is the error term.



Multiple Linear Regression

Multiple linear regression extends simple linear regression by using two or more independent variables to predict a dependent variable. The equation is

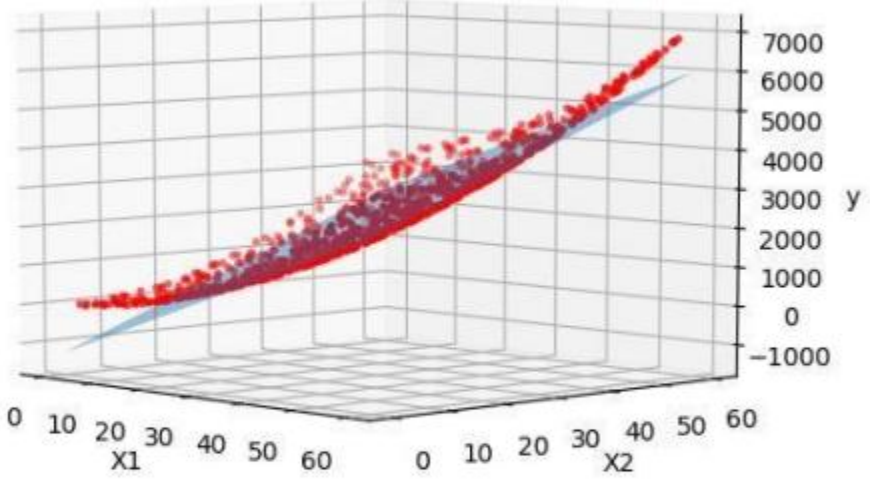
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$

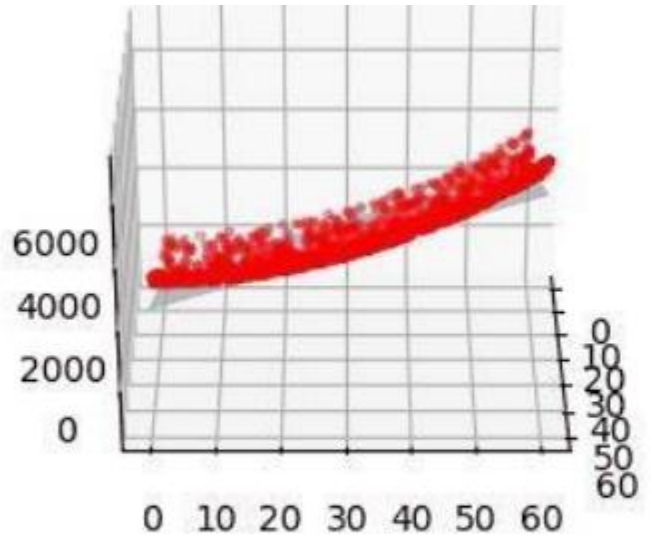
where:

- y is the predicted value (dependent variable).
- $x_1, x_2, \ldots, x_n$ are independent variables.
- β_0 is the y-intercept (constant term).
- $\beta_1, \beta_2, \ldots, \beta_n$ are coefficients for each independent variable.
- ϵ is the error term.

Multiple Linear Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$





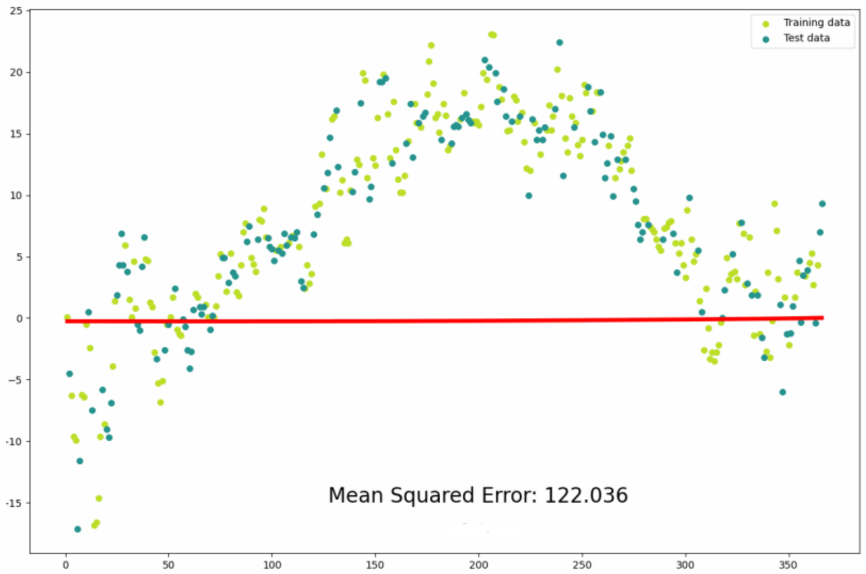
Polynomial Regression

Polynomial regression is a type of regression analysis where the relationship between the independent variable x and the dependent variable y is modeled as an n-degree polynomial. Unlike linear regression, which fits a straight line, polynomial regression fits a curve to the data.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n + \epsilon$$

where:

- y is the predicted value (dependent variable).
- x is the independent variable.
- $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are coefficients.
- n is the degree of the polynomial.
- ϵ is the error term.



Key Characteristics

1. Linear Model with Modifications

- Although polynomial regression includes non-linear terms, the overall model is still considered linear in the parameters, because the coefficients (, b1b 1 , etc.) are linear.

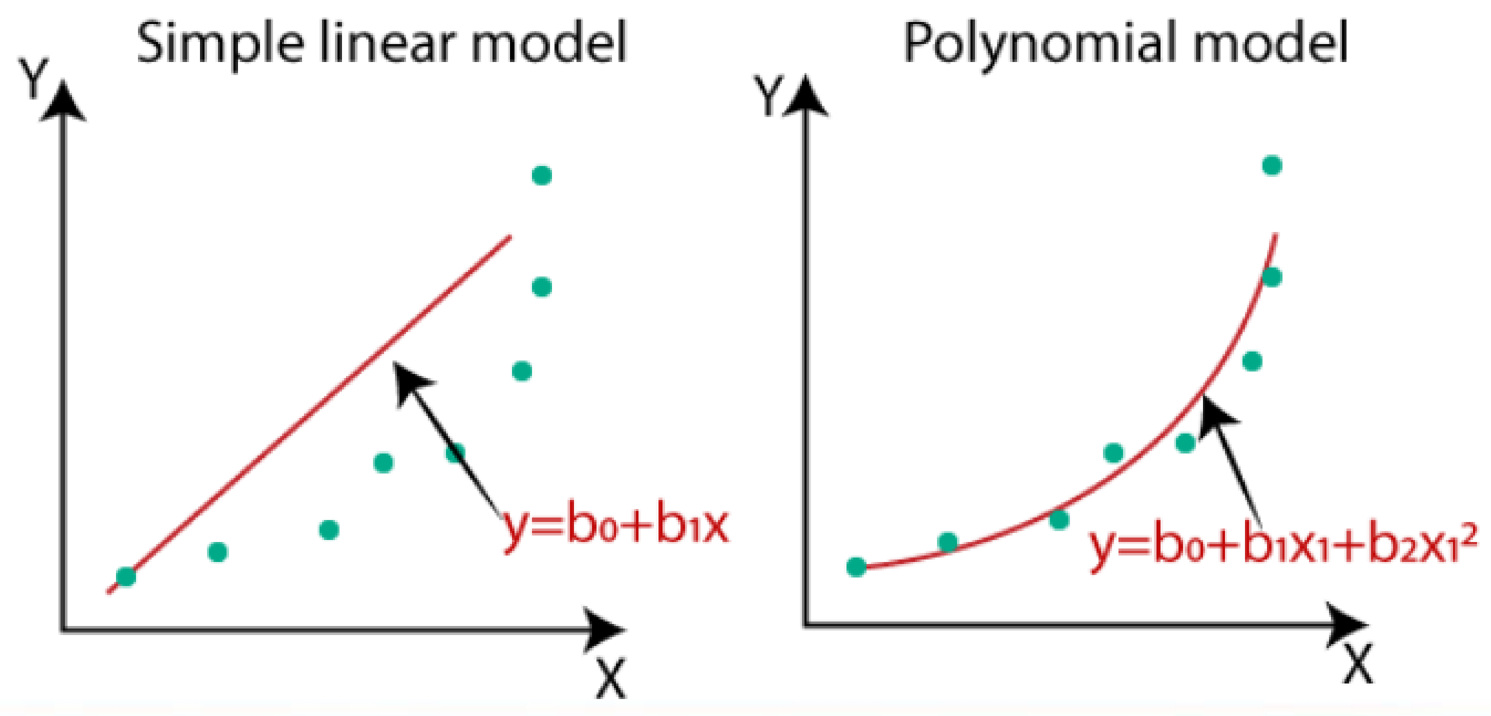
- These additional terms allow the model to **fit curves** to the data, improving accuracy when the relationship between the variables is non-linear.

2. Non-Linear Dataset

- Polynomial regression is often used for datasets where the relationship between the independent variable (X) and the dependent variable (Y) is **non-linear**.

- In the **simple linear regression** example, the model fits a straight line, but this does not always capture the true relationship. The **polynomial model** better fits the data by allowing for curves.

Key Characteristics



Cost Function

The cost function $J(\theta_0, \theta_1)$ assesses the performance of the linear regression model by measuring the difference between the predicted values and th $h_\theta(x)$ al observed values y.

Mean Squared Error (MSE):

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

- Where:
  - m represents the total number of training samples.
  - $h_\theta(x^{(i)})$ is the model's predicted value for the i-th input.
  - $y^{(i)}$ is the actual output value for the i-th input.
- **Goal**: The aim is to minimize $J(\theta_0, \theta_1)$ to enhance the model's predictive accuracy.

Different Evaluation Metrics

Mean Squared Error (MSE):

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

- Measures the average of the squared differences between actual and predicted values.
- Commonly used because it effectively penalizes larger errors and is simple to calculate.

AMIT

Different Evaluation Metrics

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

- The square root of MSE. It represents error in the same units as the original data, making it easier to interpret.

Different Evaluation Metrics

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$$

- Measures the average absolute difference between the actual and predicted values.
- Less sensitive to outliers compared to MSE, providing a more robust evaluation metric.

Gradient Descent

Purpose of Gradient Descent

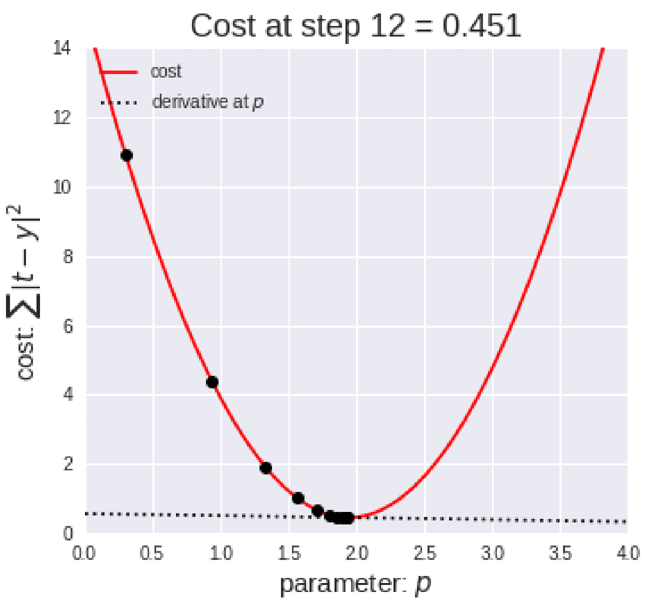
The main goal of Gradient Descent is to optimize the model parameters (θ_0 and θ_1) to minimize the cost function, which quantifies the difference between predicted and actual values. This optimization process aims to enhance the accuracy of the model.

Cost Function and Partial Derivatives

The cost function used in Gradient Descent is the Mean Squared Error (MSE represented as:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

- This function calculates the average squared difference between predicted values ($h_\theta(x)$) and actual values (y).
- To minimize this cost function, partial derivatives with respect to each parameter (θ_0 and θ_1) are computed. These derivatives show the direction and rate at which the cost function changes.



Parameter Update Rules

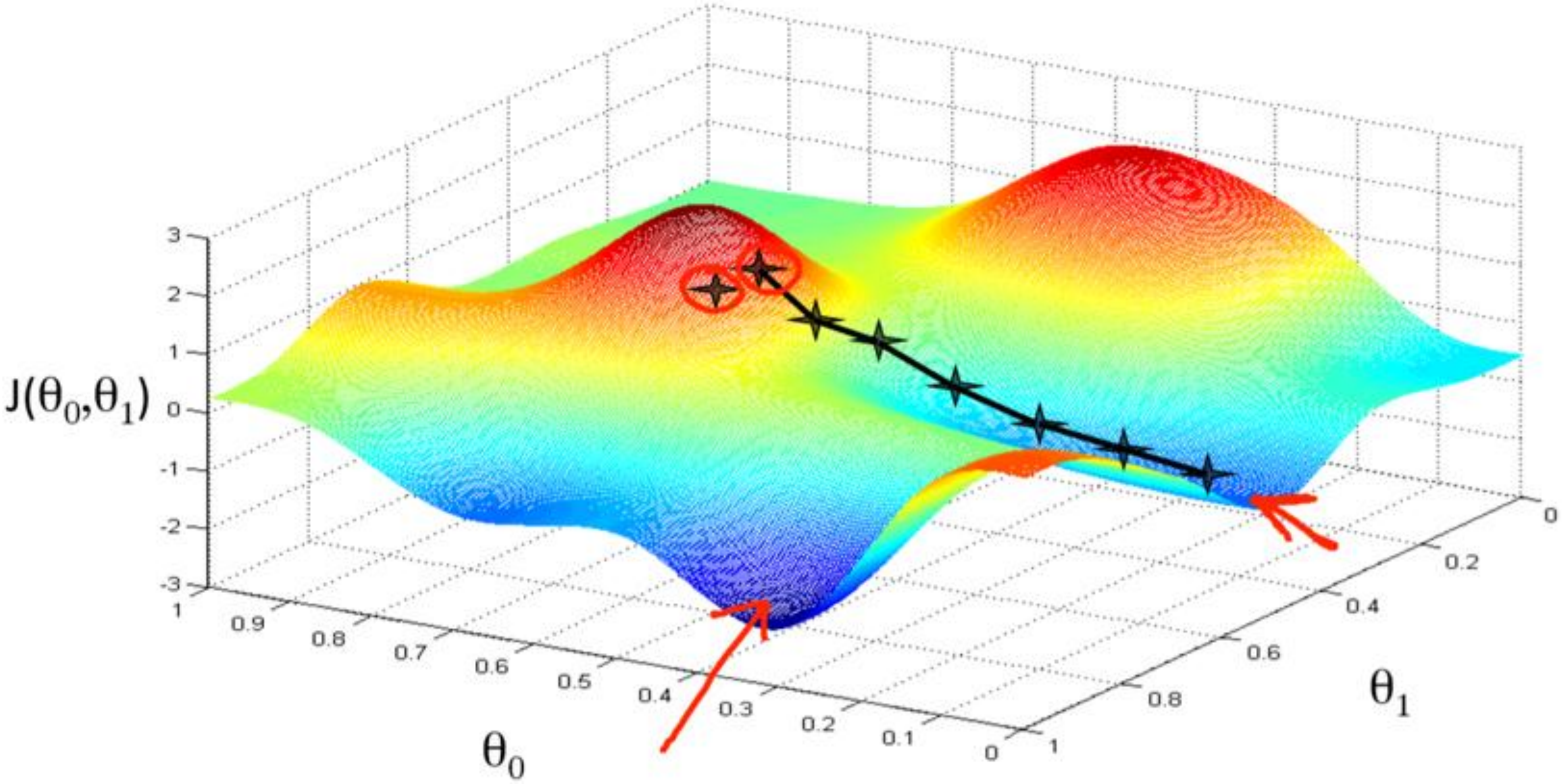
Gradient Descent iteratively adjusts the parameters using the following update rules:

$$\theta_0 = \theta_0 - \alpha \frac{\partial J}{\partial \theta_0}$$

$$\theta_1 = \theta_1 - \alpha \frac{\partial J}{\partial \theta_1}$$

Here, α represents the learning rate, a small positive value that controls the size of each step taken in the opposite direction of the gradient to reach the minimum cost.

Gradient Descent



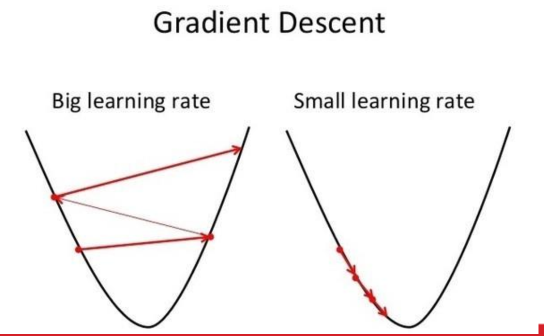
Gradient Descent Iterative Process

1. Finding Global Minimum

The objective of Gradient Descent is to achieve the global minimum of the cost function. During each iteration, parameters are updated based on the computed gradient, guiding the process toward the lowest point on the cost function curve.

2. Effect of Learning Rate (α)

- The learning rate controls the size of the steps taken toward the minimum. A small learning rate leads to slow convergence, while a large learning rate might cause overshooting or fluctuation around the minimum.
- Experimentation is usually necessary to determine an optimal learning rate that achieves stable and efficient convergence.



Quiz

What are the key differences between Linear Regression and Polynomial Regression?

Quiz

Expected Answer

- **Linear Regression**: This model assumes a linear relationship between the independent variable(s) and the dependent variable. The model fits a straight line ($y = mx + c$) through the data points, trying to minimize the error between the predicted and actual values. It works best when the relationship between the variables is approximately linear.

- **Polynomial Regression**: An extension of linear regression where the relationship between the independent variable and the dependent variable is modeled as an nth degree polynomial (e.g., $y = ax^2 + bx + c$). This allows the model to fit more complex, curved patterns in the data. It is used when the data shows a non-linear trend.