National University of Sciences & Technology

Session 2024

"COURSE TITLE"

IBM Machine Learning

Name: Warisha Masood

Discipline: SEECS, Dept NUST

Lab Name: IOT Lab

Date: 4 July,2024

Mentor/Supervisor: Dr. Rafia

1.“Supervised & Unsupervised Learning”

Supervised Learning

Supervised Learning involves teaching a machine learning model using labeled data. Labeled data means each piece of data has a label or category attached to it. Think of it like a teacher showing a student flashcards with words and pictures. The student (model) learns to recognize the words (labels) associated with each picture (data).

How do we supervise a model?

1. Teaching the Model:

We give the model lots of examples of data with their correct labels.

2. Training the Model:

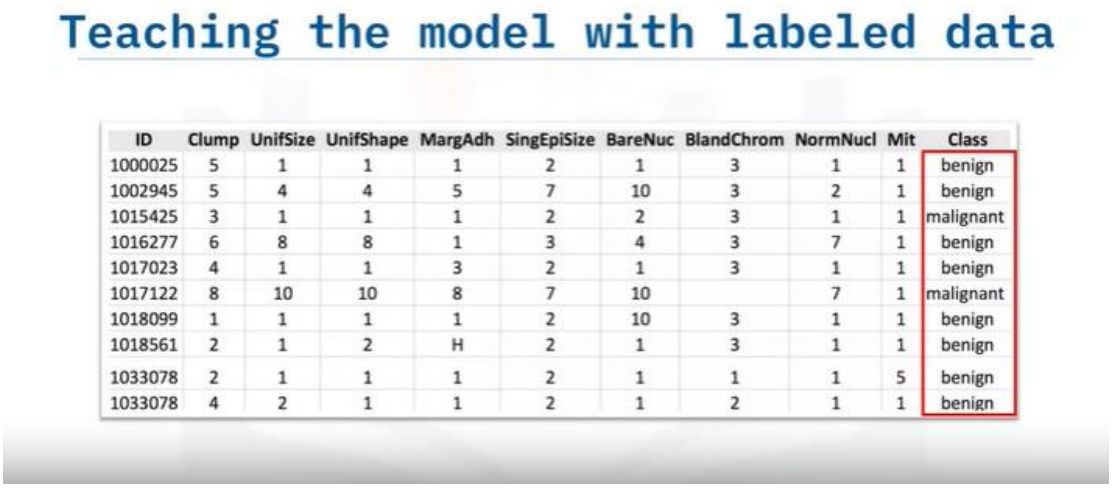
The model uses these examples to learn patterns and make predictions about new, unseen data.

Example:

- If you have medical data about patients, the labels might be 'benign' or 'malignant' (types of tumors).

- Each patient’s data includes features like clump thickness, cell size, etc.

- The model learns from this labeled data to predict if a new patient's tumor is benign or malignant.



Types of Supervised Learning:

1. Classification:

Predicting categories. For example, predicting if an email is 'spam' or 'not spam.'

2. Regression:

Predicting continuous values. For example, predicting the price of a house based on its size, location, etc.

Unsupervised Learning

Unsupervised Learning means letting the model find patterns in data without any labels. The model independently identifies structures and relationships in the data, like exploring a new city without a map.

How does it work?

1. No Labels:

We give the model data without any predefined labels.

2. Finding Patterns:

The model looks for patterns and groupings in the data on its own.

Example:

Clustering Customers Based on Purchase Behavior

Imagine you run an online store and have data on your customers' purchasing behavior. This data includes information such as:

- How often they buy
- The types of products they purchase
- The amount they spend

You want to group your customers into segments to better target your marketing efforts, but you don't have predefined labels (like "frequent buyers" or "bargain hunters"). Here's how unsupervised learning can help:

Step-by-Step Example:

1. **Collect Data:**
    - Gather data on various customer behaviors. For example:
        - Customer A: Buys frequently, spends a lot.

        - Customer B: Buys occasionally, spends moderately.
        - Customer C: Buys rarely, spends a little.
2. **Choose an Algorithm:**
    - Use a clustering algorithm like K-means clustering. This algorithm will group customers into clusters based on their purchasing behavior.
3. **Process the Data:**
    - The algorithm will analyze the data and find patterns.
    - It will group customers with similar behaviors into clusters.
4. **Interpret the Clusters:**
    - After clustering, you might get results like:
        - Cluster 1: Frequent buyers who spend a lot.
        - Cluster 2: Occasional buyers who spend moderately.
        - Cluster 3: Rare buyers who spend little.
5. **Apply the Insights:**
    - Use these clusters to tailor your marketing strategies:
        - Offer special deals to frequent buyers to maintain their loyalty.
        - Send occasional buyer's reminders and promotions to encourage more frequent purchases.
        - Offer discounts to rare buyers to increase their spending.
        - 

- Red Cluster (Cluster 1): Frequent buyers who spend a lot.
- Green Cluster (Cluster 2): Occasional buyers who spend moderately.
- Blue Cluster (Cluster 3): Rare buyers who spend little.
- Yellow Stars: Centroids of each cluster, representing the center of each group.

The x-axis represents the frequency of purchase, and the y-axis represents the amount spent. The clustering algorithm has grouped the customers into three clusters based on their purchasing behavior.

Types of Unsupervised Learning:

1. Clustering:

Grouping similar data points together. For example, grouping customers with similar buying habits.

2. Dimensionality Reduction:

Simplifying data by reducing the number of features while keeping important information.

3. Market Basket Analysis:

Finding associations between items. For example, if someone buys bread, they are likely to buy butter too.

4. Density Estimation:

Estimating the data distribution to understand the data structure better.

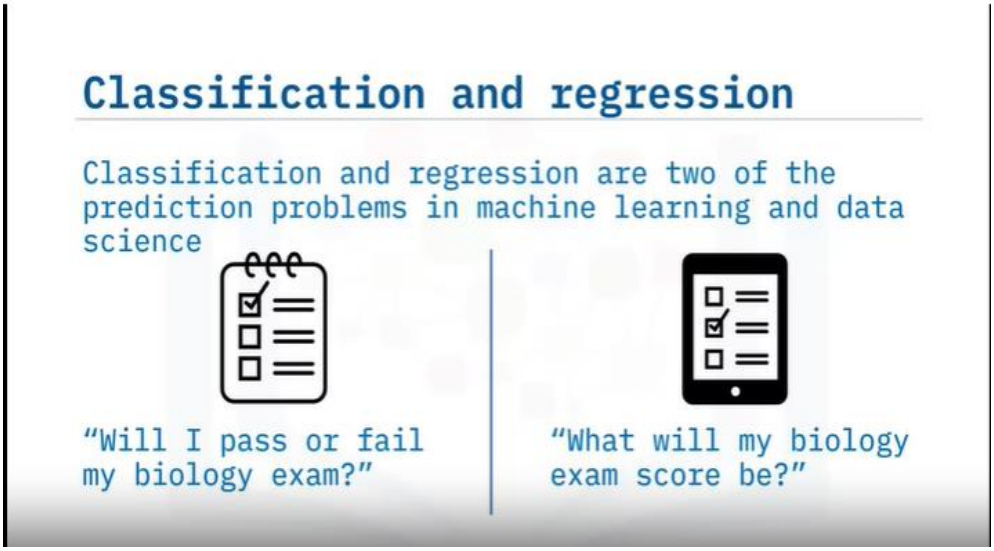
Key Differences

- Supervised Learning uses labeled data and includes techniques like classification and regression.

- Unsupervised Learning uses unlabeled data and includes techniques like clustering and market basket analysis.

2.“Classification and Regression”

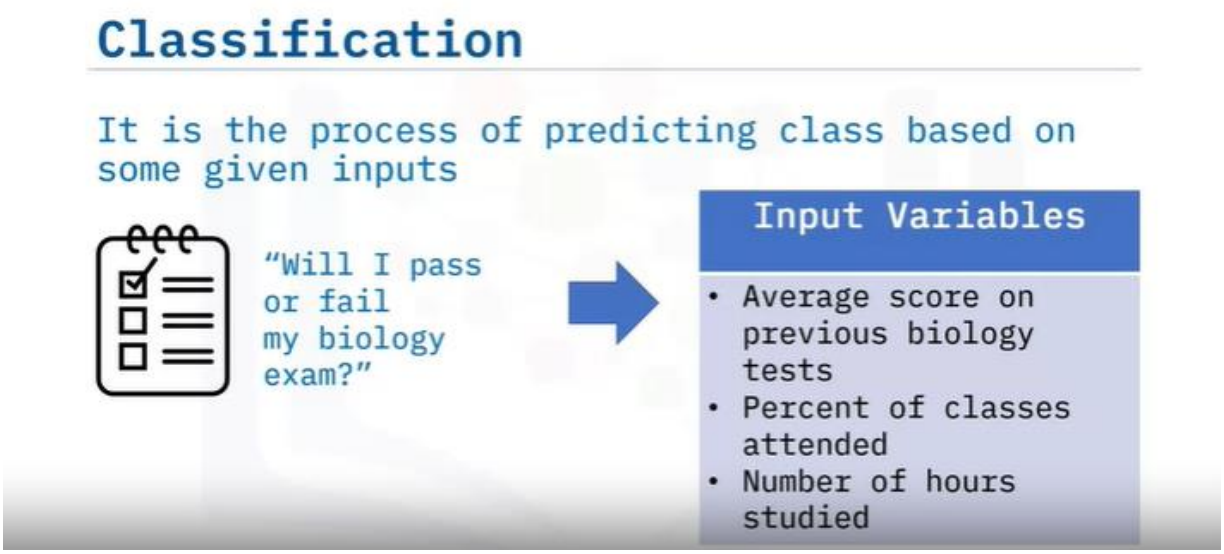
Classification and regression are two types of prediction problems in machine learning and data science.



Classification

Classification is the process of predicting which category or class an input belongs to.

Example: "Will I pass or fail my biology exam?"



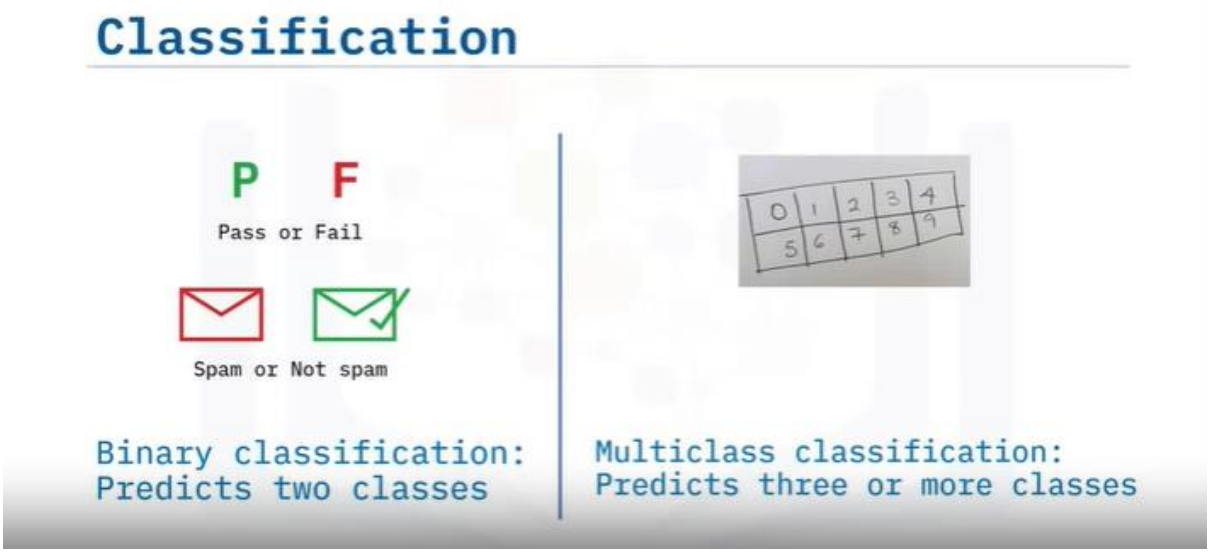
-Inputs (Features):

Data is used to make predictions. Examples include:

- Average score on previous tests
- Percent of classes attended
- Number of hours studied

- ***Binary Classification:*** Predicts one of two outcomes. Example: Pass or fail.

- ***Multiclass Classification:*** Predicts one of three or more outcomes. Example: Predicting the fruit type (apple, orange, mango).



Common Algorithms:

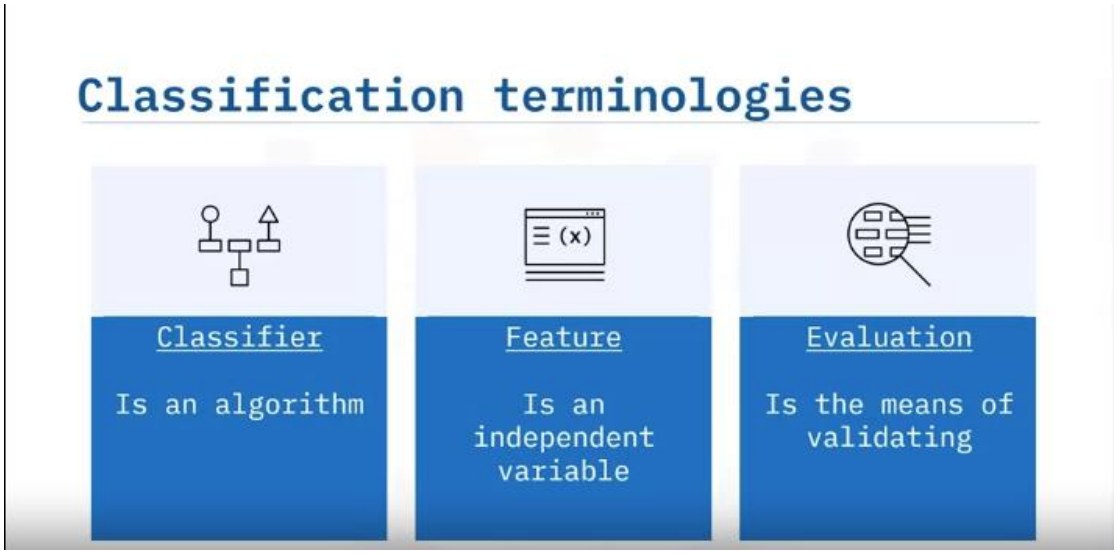
- **K-Nearest Neighbors (KNN):** Classifies a data point based on the majority class of its nearest neighbors. Example: Classifying a new fruit based on the closest fruits in a dataset.
- **Logistic Regression:** Predicts the probability of a binary outcome. Example: Predicting whether a student will pass or fail an exam based on study hours and previous scores.
- **Decision Trees:** Uses a tree structure to make decisions based on feature values. Example: Classifying emails as spam or not spam based on features like word presence and email length.

Important Terminology

- ***Classifier***: A machine learning algorithm used to solve classification problems.

- ***Feature:*** An independent variable used as input in the model.

- ***Evaluation:*** The process of validating how well a model performs.



Types of Learners

- ***Lazy Learner:*** Doesn't generalize during training. Example: KNN.

- ***Eager Learner:*** Spends time training and generalizing the model. Example: Decision Trees, Logistic Regression.

Regression

Regression is the process of predicting a continuous value based on input features.

Example: "What will my biology exam score be?"

- ***Inputs (Features):***

Data used to make predictions. Examples include:

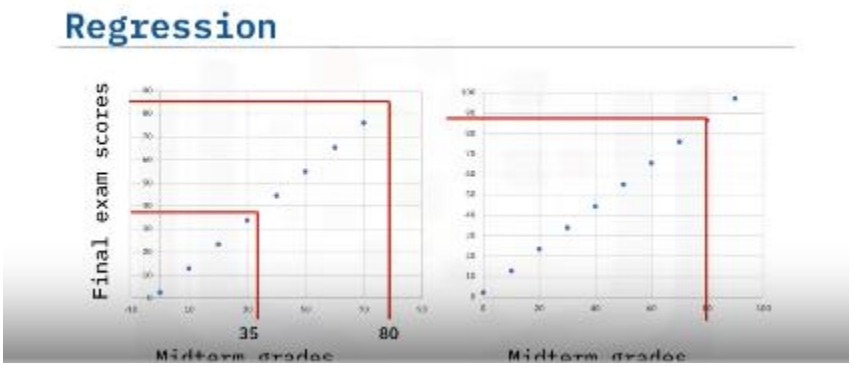
- Hours of sleep
- Hours studied

Regression Analysis Simplified

Regression is the relationship between a dependent variable (outcome) and an independent variable (predictor). It helps predict continuous outcomes.

Example:

Predicting Final Exam Scores



Calculation Explained:

Line of Best Fit:

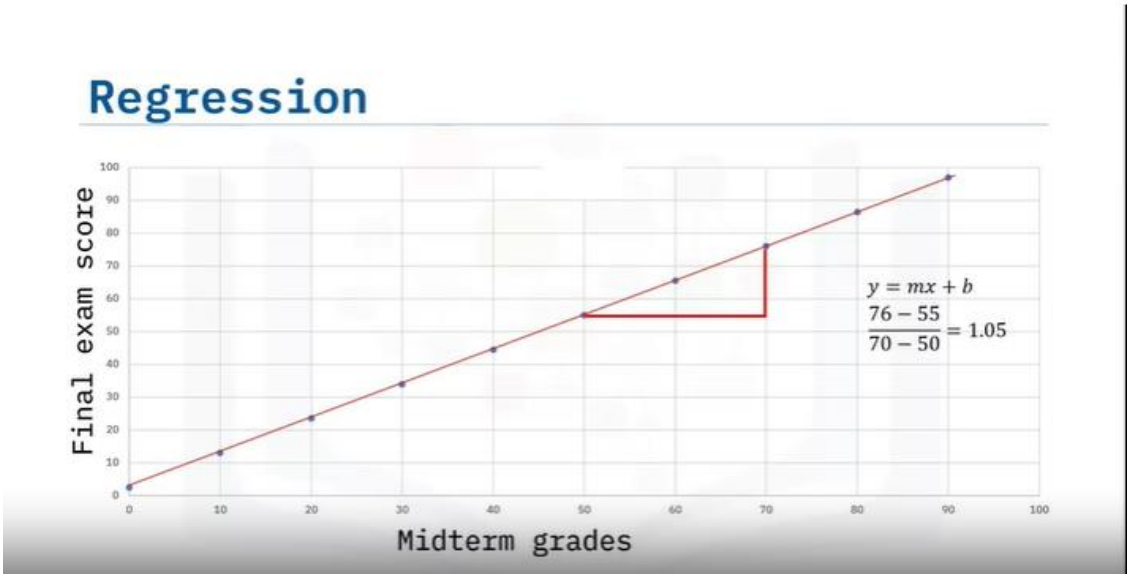
- Purpose: To make educated guesses for values within or outside the dataset.
  - Equation: $y=mx+by$
  - Slope (rise over run)
  - b: Y-intercept (value of y when x is 0)

Slope (m):

The slope indicates how much the dependent variable (final grade) changes for every one-unit change in the independent variable (midterm grade).

Example Calculation:

- Points Chosen: (70, 76) and (50, 55)
- Formula: slope (m)=riserun=76−55/70−50=2120=1.05\
- Interpretation:
- Meaning: For every 1-mark increase in the midterm grade, the final exam score increases by 1.05 marks.
- Example: If a student's midterm score increases from 50 to 51, their final score is predicted to increase by 1.05 marks.



Y-Intercept (b):

The y-intercept is the point where the line crosses the y-axis. It represents the value of the dependent variable (final grade) when the independent variable (midterm grade) is 0.

Example Calculation:

- Equation: y=mx+by
- Given Slope (m): 1.05
- Given Point: (50, 55)

Find b:

Substitute one point into the equation
55=1.05×50+b

$$55=52.5+b$$

$$b=55-52.5=2.5$$

Interpretation:

- Meaning: If a student scores 0 on the midterm, their predicted final score is 2.5.
- Example: Even without any midterm marks, the model predicts the student would score 2.5 on the final exam due to other factors not captured by the midterm grade.

Types of Regression

Linear Regression:

Predicts outcomes using a straight line.

Example: Predicting house prices or exam scores.

Multiple Regression:

Uses multiple independent variables.

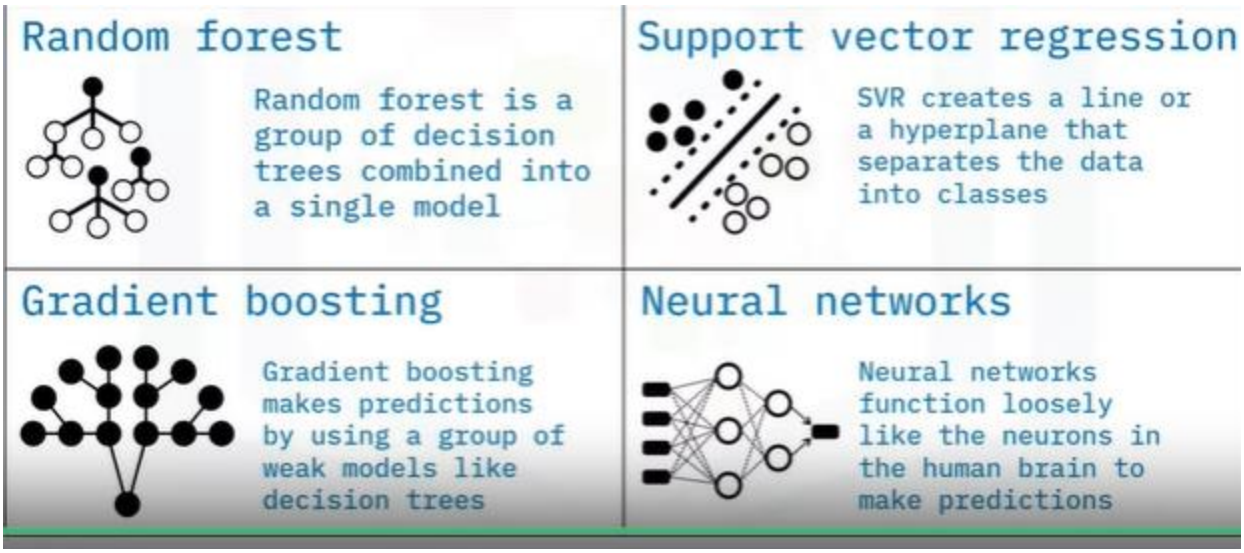
Example: Predicting final scores based on midterm scores and attendance percentage.

Polynomial Regression:

Accounts for non-linear relationships using polynomials.

Example: Complex relationships not captured by a straight line.

Regression Algorithms:



"Evaluating Machine Learning Model"

1. Splitting Data

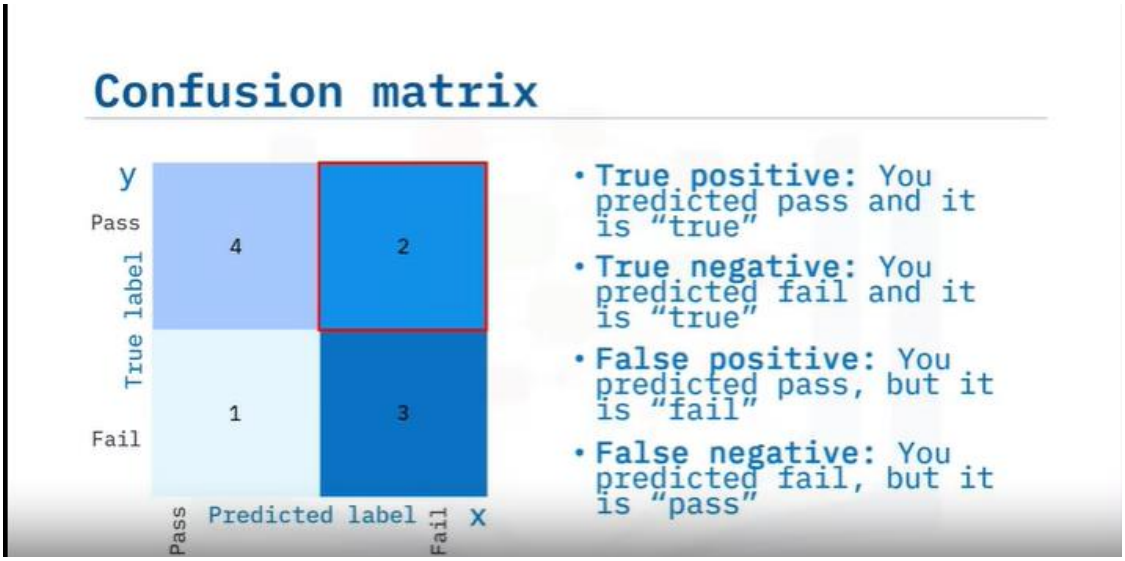
- Training Set: Used to teach the model with lots of examples.
- Test Set: Used to check how well the model performs on new data.

2. Accuracy

- The percentage of correct predictions made by the model.
- Example: If the model predicts "Pass" or "Fail" on a test and gets 7 out of 10 correct, the accuracy is 70%.

3. Confusion Matrix

- A table that shows how well the model's predictions match the actual values.
- Components:
    - True Positive (TP): Correctly predicted "Pass" and it was a pass.
    - True Negative (TN): Correctly predicted "Fail" and it was a fail.
    - False Positive (FP): Predicted "Pass" but it was a fail.
    - False Negative (FN): Predicted "Fail" but it was a pass.



4. Precision

- Measures how many of the predicted positives are correct.
- Formula: Precision = TP / (TP + FP)
- Example: If 4 out of 5 predicted "Pass" were correct, precision is 80%.

5. Recall

- Measures how many of the actual positives were correctly predicted.
- Formula: Recall = TP / (TP + FN)
- Example: If 4 out of 6 actual "Pass" were correctly predicted, recall is about 66.7%.

6. F1-Score

- Combines precision and recall into one score.

7. Regression Model Evaluation

1. Mean Squared Error (MSE):

Average of the squared differences between predicted and actual values.

2. Root Mean Squared Error (RMSE):

Square root of MSE.

3. Mean Absolute Error (MAE):

Average of the absolute differences between predicted and actual values.

4. R-Squared:

Shows how well the model fits the data.

"Deep Learning"

Deep Learning:

A type of machine learning that uses models similar to the human brain to learn from data.

Purpose: It automates finding important features from data, like recognizing objects in photos.

Key Concepts

1. Artificial Neural Network (ANN):

A network of connected nodes (neurons) that mimic the brain.

Components:

- Neurons: Compute and process data.

Layers:

- Input Layer
- Hidden Layers
- Output Layer

2. Activation Function:

A function that determines if a neuron should be activated or not.

-Purpose: Normalizes outputs to a range (e.g., 0 to 1).

3. Deep Neural Network:

A neural network with multiple hidden layers.

Feature: Allows for more complex learning and predictions.

How It Works:

1. Input Layer: Takes in data (e.g., number of rooms in a house).

2. Hidden Layers: Assign weights and process data.

3. Output Layer: Gives the prediction (e.g., house price).

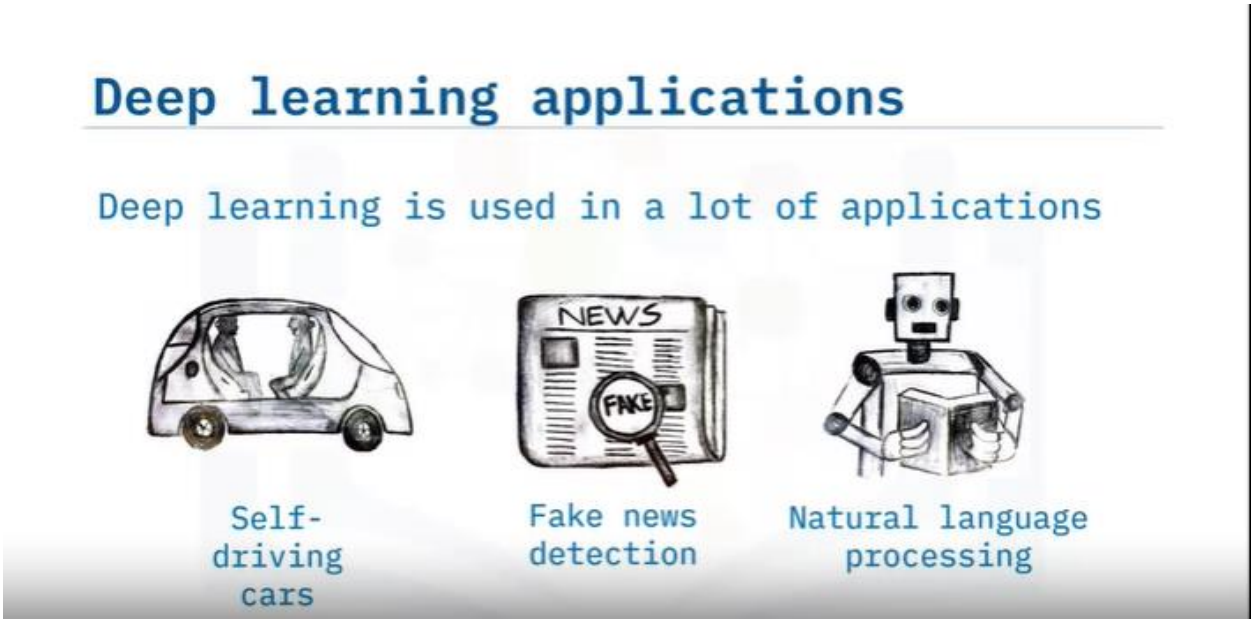
Applications of Deep Learning

1. Self-Driving Cars

2. Fake News Detection

3. Natural Language Processing

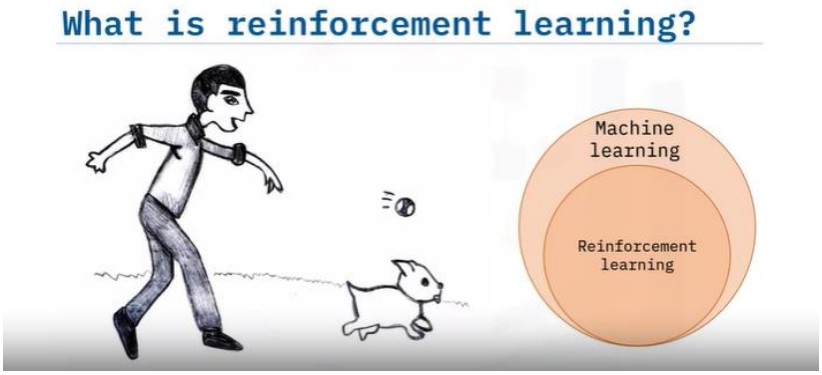
4. Game Playing



“Reinforcement”

Reinforcement Learning:

A type of machine learning where a system learns by getting rewards for correct actions and penalties for mistakes, like training a dog with treats.



Key Concepts

1. Agent:

The learner or decision-maker. Think of it like the dog playing fetch.

2. Environment:

Everything the agent interacts with, like the park where the dog plays.

3. Action:

What the agent decides to do, like the dog catching or missing the ball.

4. State:

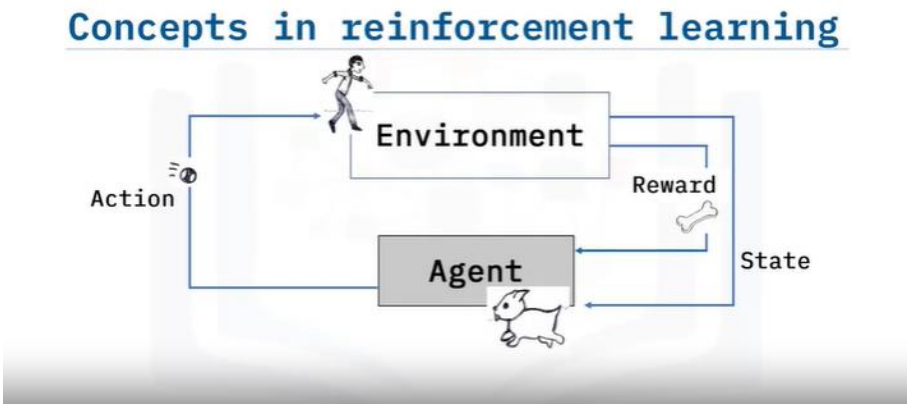
The current situation or position of the agent.

5. Reward:

The feedback given to the agent after an action.

6. Policy:

A strategy for deciding what actions to take in different states.



Types of Reinforcement

1. Positive Reinforcement:

Giving rewards to increase the likelihood of a behavior.

Example: Treats for catching the ball.

2. Negative Reinforcement:

Strengthening behavior by removing something negative.

Example: Stopping a loud noise when the correct action is taken.

3. Neutral Reinforcement:

No reward or penalty, so the behavior doesn't change.

Example: Ignoring the dog chewing shoes.

Applications of Reinforcement Learning

- Fleet Management:
- Example: Deciding where to send delivery trucks for maximum profit.
- Gaming:
- Example: Teaching AI to play and win at chess.
- Robotics:
- Example: Training a robot to pick and place objects correctly

“AI & Gen AI”

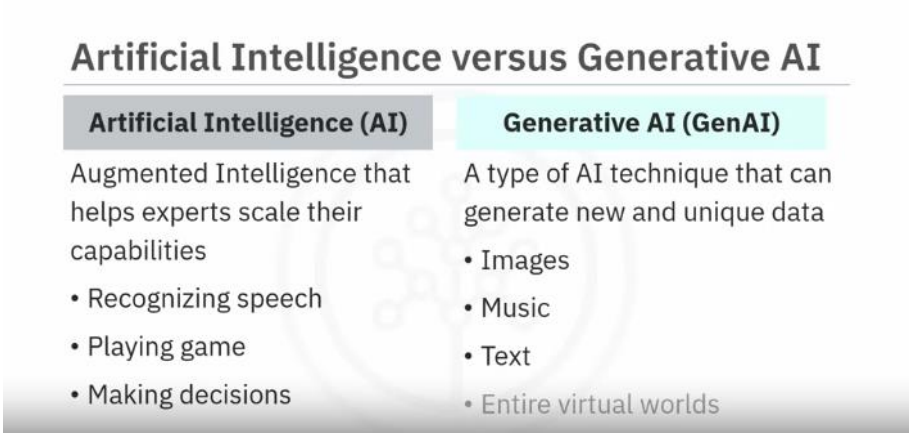
What is Artificial Intelligence (AI)?

AI helps experts by doing tasks like understanding speech, playing games, and making decisions. -

Example Tasks: AI can help a phone understand your voice commands or make a robot play chess.

What is Generative AI (Gen AI)?

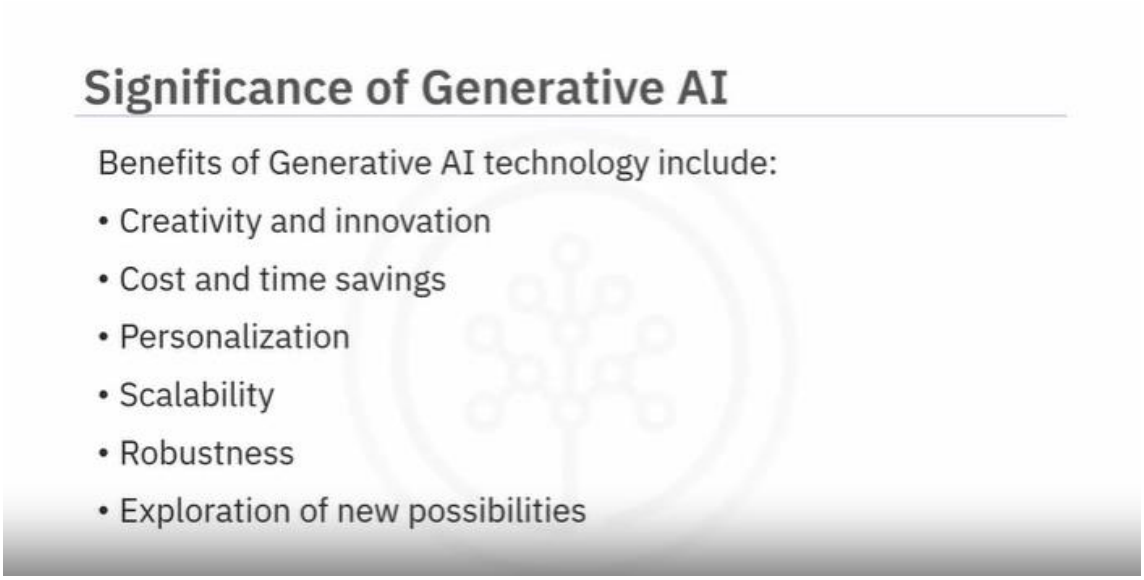
Gen AI can create new things like pictures, music, text, and even virtual worlds.



Large Language Models (LLM)

These are AI models that understand and create human language.

Significance Of Generative AI:



Gen AI Applications:

Area	What Generative AI Does
Healthcare	Creates medical images and predicts treatments.
Agriculture	Helps farmers grow better crops.
Biotechnology	Assists in creating new medicines and therapies.
Forensics	Analyzes evidence to help solve crimes.
Environmental Conservation	Protects endangered animals.
Creative Fields	Produces new digital art, music, and videos.
Gaming	Generates new game levels and characters.
Fashion	Designs clothes and suggests outfits.
Robotics	Teaches robots new movements.
Education	Creates personalized learning materials.
Data Augmentation	Generates additional data to train other AI models.