Machine Learning Fundamentals **Eman Shahid**

**Introduction:**

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that enables systems to learn from data and improve performance without being explicitly programmed. It is widely used in various domains, including healthcare, finance, natural language processing, computer vision, and recommendation systems

**Applications of Machine Learning:**

- Healthcare: disease prediction, drug discovery  
- Finance: fraud detection, stock price prediction  
- Natural Language Processing (NLP): chatbots, sentiment analysis  
- Computer Vision: image classification, object detection  
- E-commerce: recommendation systems, personalization

**Types of Machine Learning:**

**Supervised Learning:**

Supervised Learning uses labeled datasets to train algorithms. The model learns the mapping between inputs (features) and outputs (labels). Examples include regression and classification tasks.

**Unsupervised Learning:**

Unsupervised Learning uses unlabeled data to find hidden patterns or groupings. It is mainly used for clustering and dimensionality reduction.

**Reinforcement Learning:**

Reinforcement Learning trains agents to make decisions by interacting with an environment. The agent receives rewards or penalties based on its actions and learns optimal policies over time.

**Supervised Learning Models:**

**Linear Regression:**

A regression technique that models the relationship between independent variables (features) and a dependent variable (target) using a straight line.

**Logistic Regression:**

A classification technique that estimates the probability of a binary outcome using a logistic function.

**K-nearest neighbor:**

A non-parametric method for classification and regression. It predicts outcomes based on the majority class or average of the k-nearest data points.

**Decision Tree:**

A hierarchical model that splits data into subsets using feature-based conditions. Useful for both regression and classification.

**Random Forest:**

An ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting.

**Native Bayes:**

A probabilistic classifier based on Bayes’ theorem, assuming independence among predictors. Efficient for text classification and spam filtering.

**SVM:**

A classification technique that finds an optimal hyperplane that separates data into different classes with maximum margin.

**Unsupervised Learning Models:**

**K means Clustering:**

A partition-based clustering algorithm that groups data into k clusters based on similarity.

**PCA:**

A dimensionality reduction technique that transforms correlated features into uncorrelated principal components.

**DBSCAN:**

A density-based clustering algorithm that groups together closely packed points and identifies outliers as noise.

**Hierarchical Clustering:**

A clustering method that builds a tree of clusters using agglomerative or divisive approaches.

**Gaussian Mixture Model:**

A probabilistic model that represents data as a mixture of multiple Gaussian distributions.

**Reinforcement learning Model:**

Reinforcement Learning (RL) is based on an agent-environment interaction framework. The agent learns to maximize cumulative rewards by exploring and exploiting actions. Popular RL algorithms include Q-Learning and Deep Q-Networks (DQN).

**Summary Table of Ml model:**

Supervised Models: Linear Regression, Logistic Regression, KNN, Decision Trees, Random Forest, Naive Bayes, SVM  
Unsupervised Models: K-Means, PCA, DBSCAN, Hierarchical Clustering, GMM  
Reinforcement Learning: Q-Learning, DQN