ACM Machine Learning Question and Answer

Q1) What is Machine Learning?

Machine learning (ML) is a subdomain of data science and ai that relies on the data to enable machines to exhibit human intelligence, to train models to reason and work independently without much human supervision.

Types of Machine Learning

Supervised Learning: The model is trained on labelled data, and output is given for the input and the model is expected to find reasoning between the pairs is called supervised learning.

Example: Spam email filtering.

Unsupervised Learning: The model is trained on unlabelled data and discovers patterns from the data. There is no inherent understanding of the output in unsupervised learning

Application: Marketing customer segmentation.

Reinforcement Learning: It is a type of learning where we have a system that constantly changes the structure of the model to better fit the task the closest it is to how humans function The model learns by interacting in an environment, being rewarded for correct actions and penalized for incorrect actions.

Application: Training autonomous vehicle robots.

Q2)

1. Linear Regression

the main objective being to predict the next value so works for predicting the next value

- **Use Case**: Predicting house prices based on features like size, location, etc.

2. Decision Trees

A tree-like structure where each node represents a decision based on input features to mimic and recreate a mind map of the application at hand

- **Use Case**: Medical diagnosis (e.g., classifying diseases based on symptoms).

3. Random Forest

When there a lot of tresses working to make an answer and the final verdict is drawn based based on the most common decision

- **Use Case**: Fraud detection in banking.

4. Support Vector Machines (SVM)

When the data is messy in the lower dimensions we try to put it in higher dimensions with the help of kernels where it might be easy to make the distinction between the data points for example it is difficult to make equation of circle in 2d but if the data were to be in parabolic shape and we were to use a plane to cut the graph would be much easier

- **Use Case**: Face recognition systems.

5. k-Nearest Neighbours (k-NN)

it is a learning algorithm where we try to find the centre of points of the cluster of data to make a classification of data points to see where other data points fall so to assume on how similar to the data points

- **Use Case**: Recommender systems (e.g., movie recommendations).

6. Neural Networks

NN are attempting to recreate the brain structure to better reason and learn the surroundings with simple weight and biases to make classification as well as regression

- **Use Case**: Image recognition in self-driving cars.

Q3) **Comparing Models**

1. Linear Regression

Advantages: Simple, interpretable.

Disadvantages: Sensitive few parameter.

Usecase : structured data

2. Decision Trees

Advantages: simple understanding, handling multiple outcomes, and non-linearity well.

Disadvantages: Prone to overfitting.

Usecase : structured data

3. Random Forest

Advantages: less overfitting, high accuracy.

Disadvantages: Slower for large datasets.

Usecase : structured data

4. Support Vector Machines (SVM)

Advantages: Works well in high dimensions.

Disadvantages: Computationally expensive.

Usecase : structured data

5. k-Nearest Neighbors (k-NN)

Advantages: direct implementation and simple.

Disadvantages: Computationally expensive.

Usecase : structured data

6. Neural Networks

Advantages: Handles complex patterns.

Disadvantages: needs large data set and little to very complex understanding of its working.

Usecase : Unstructured data

Q4)

1. Accuracy:

how well the answers fit with the test data overall meaning a number of bets hit out of the total hits

Use case: When false positives and false negatives are equally important.

2. Precision:

Precision is when see how close the given answer to the right answer

Use case: When false positives are costly (e.g., spam detection).

3. Recall:

in a binary classification how many correct positive answers were given as getting a false answer is very costly.

Use case: When false negatives are costly (e.g., cancer detection).

4. F1 Score:

A simple harmonic of recall and precision to deal with imbalanced or when we don't have proper data set at hand (large data set of positive data but not so much of negative data in a diagnosis)

Use case: When there is an imbalance in class distribution.