Machine Learning Questions

Question 1

Explain the difference between supervised and unsupervised machine learning?

SUPERVISED MACHINE LEARNING

- Supervised learning is a machine learning method in which models are trained using labelled data
- In Supervised learning, models need to find the mapping function to map the input variable with the output variable
- It needs supervision to train the model
- This is can be used for two types of problems: Classification and Regression

UNSUPERVISED MACHINE LEARNING

- Unsupervised learning is a machine learning method in which patterns inferred from the unlabelled input data
- The goal of unsupervised learning is to find the structure and patterns from the input data
- It does not need any supervision, Instead it finds patterns from the data by its own
- This can be used for two types of problems:
  - Clustering and Association

Question 2

Explain the difference between KNN and K-Means clustering?

K-NEAREST NEIGHBOR

- K-NN is a Supervised machine learning
- K-NN is a Classification or Regression algorithm
- K-NN is a lazy learner (A lazy learner does not have a model fitting that means does not have a training phase)

K-MEANS CLUSTERING

- K-means is an Unsupervised machine learning
- K-means is a clustering machine learning algorithm
- K-means is an eager learner (An eager learner has a model fitting that means a training step)

Question 3

What is the difference between Classification and Regression?

REGRESSION

- Regression finds correlations between dependent and independent variables
- It help predict continuous variables such as house price, market trends, etc
- Types of Regression:

  Decision Tree Regression, Principal Components Regression, Polynomial Regression, Random Forest Regression, Simple linear, Support Vector Regression

CLASSIFICATION

- Classification finds functions that help divide the dataset into classes based on various parameters
- When using it, a program gets taught on the training dataset and categorizes the data into various categories depending on what it learned
- Used for things like Email and Spam classification, identifying cancer tumour cells, etc
- Types of Classification:

  Decision Tree Classification, K-Nearest Neighbors, Logistic Regression, Naïve Bayes, Random Forest Classification, Support Vector Machines

Question 4

How will you make a 5th class student to understand what a unsupervised Machine Learning problem will be like?

- Suppose your friend invites you to his party where you meet total strangers. Since you have no idea about them, you will mentally classify them on the basis of gender, age group, dressing, etc.
- In this scene, the strangers represent unlabelled data and the process of classifying unlabelled data points in nothing but unsupervised learning
- Since you didn't use any prior knowledge about the people and the features, You classified them in the instant you saw them.
- This becomes an unsupervised Machine Learning Problem.

Question 5

Explain the 4 terms in the confusion matrix?

- **True Positive (TP) :** If the actual value and predicted values are true the it is True Positive
- **False Positive (FP) :** If the predicted is value is False but the actual value is True then it is False Positive
- **True Negative (TN) :** If the actual and predicted value are False then it is True Negative
- **False Negative (FN) :** If the predicted value is True but the actual value is False then it is False Negative

Question 6

Can youcome with the advantages of Naïve Bayes?

- It is simple and easy to implement
- It doesn’t require as much training data
- It handles both continuous and discrete data
- It is highly scalable with the number of predictors and data points
- It is fast and can be used to make real-time predictions
- It is not sensitive to irrelevant features
- If the conditional Independence assumption holds, it could give great results

Question 7

Explain Ensemble Learning

- In simple English, ensemble refers to a group of items
- An ensemble method is a technique which uses multiple independent similar or different models/weak learners to derive an output or make some predictions
- An ensemble can also be built with a combination of different models like random forest, SVM, Logistic regression etc.
- There are two widely used methods of Ensemble techniques. They are :

  Bagging and Boosting
- Ensemble learning is used for 2 main reasons : Performance and Robustness

Question 8

Explain Dimensionality Reduction in Machine Learning

- Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principal variables.
- The various methods used for Dimensionality Reduction include:
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - Generalized Discriminant Analysis (GDA)
- It can be divided into feature selection and feature extraction.

Question 9

What should you do when your model is suffering from low bias and high variance?

- Firstly Low bias and High variance refers to overfitting condition i.e., the model performs very well with the training data but gives disappointment results when done with the training data
- In such situations, the methods to solve are :
  - Bagging Technique (Random Forest Algorithm) to tackle high variance problem
  - Boosting Technique
  - Regularization
  - Cross-Validation method
  - Remove low co-related features
  - Train with more data (Data Augmentation)

Conclusion

Great! You havemade 9 Questions

Keep Learning