Machine learning 14

1. What is the concept of supervised learning? What is the significance of the name?

Ans:-- Supervised learning **uses a training set to teach models to yield the desired output**. This training dataset includes inputs and correct outputs, which allow the model to learn over time. The algorithm measures its accuracy through the loss function, adjusting until the error has been sufficiently minimized.

2. In the hospital sector, offer an example of supervised learning.

Ans:- In both these cases, the computer is approximating what a trained physician is already capable of doing with high accuracy. **Supervised learning is often used to estimate risk**.

3. Give three supervised learning examples.

Ans:- One practical example of supervised learning problems is **predicting house prices**. How is this achieved? First, we need data about the houses: square footage, number of rooms, features, whether a house has a garden or not, and so on. We then need to know the prices of these houses, i.e. the corresponding labels.

4. In supervised learning, what are classification and regression?

Ans:- **Classification is the task of predicting a discrete class label.** **Regression is the task of predicting a continuous quantity**.

5. Give some popular classification algorithms as examples.

Ans:- The best example of an ML classification algorithm is **Email Spam Detector**. The main goal of the Classification algorithm is to identify the category of a given dataset, and these algorithms are mainly used to predict the output for the categorical data.

6. Briefly describe the SVM model.

Ans:- SVM or Support Vector Machine is **a linear model for classification and regression problems**. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes.

7. In SVM, what is the cost of misclassification?

Ans:- Misclassification costs allow you to specify the relative importance of different kinds of prediction errors. Misclassification costs are basically **weights applied to specific outcomes**. These weights are factored into the model and may actually change the prediction (as a way of protecting against costly mistakes).

8. In the SVM model, define Support Vectors.

Ans:- **The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane** are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

9. In the SVM model, define the kernel.

Ans:- “Kernel” is used due to **a set of mathematical functions used in Support Vector Machine providing the window to manipulate the data**. So, Kernel Function generally transforms the training set of data so that a non-linear decision surface is able to transform to a linear equation in a higher number of dimension spaces.

10. What are the factors that influence SVM's effectiveness?

Ans:- The effectiveness of SVM depends on **the selection of kernel, kernel's parameters and soft margin parameter C**. . Each pair of parameters is checked using cross validation, and the parameters with best cross validation accuracy are picked.

11. What are the benefits of using the SVM model?

Ans:- SVM model has some excellent advantages such as (a) **highly effective** (b) efficient memory (c) works well with both semi-structured and unstructured data (d) effective for high dimensional data (e) less overfitting issues (f) works well for small dataset [27], [28].

12. What are the drawbacks of using the SVM model?

Ans:- **It does not execute very well when the data set has more sound i.e. target classes are overlapping**. In cases where the number of properties for each data point outstrips the number of training data specimens, the support vector machine will underperform.

14. What are some of the benefits of the kNN algorithm?

Ans:- **Can learn non-linear decision boundaries when used for classfication and regression**. Can came up with a highly flexible decision boundary adjusting the value of K.

15. What are some of the kNN algorithm's drawbacks?

Ans:-

It's main disadvantages are that **it is quite computationally inefficient and its difficult to pick the “correct” value of K**. However, the advantages of this algorithm is that it is versatile to different calculations of proximity, it's very intuitive and that it's a memory based approach.

16. Explain the decision tree algorithm in a few words.

Ans:- A decision tree is **a type of supervised machine learning used to categorize or make predictions based on how a previous set of questions were answered**. The model is a form of supervised learning, meaning that the model is trained and tested on a set of data that contains the desired categorization.

17. What is the difference between a node and a leaf in a decision tree?

Ans:- The root node is just the topmost decision node. In other words, it is where you start traversing the classification tree. **The leaf nodes (green), also called terminal nodes, are nodes that don't split into more nodes**. Leaf nodes are where classes are assigned by majority vote.

18. What is a decision tree's entropy?

Ans:- Entropy. A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogenous). **ID3 algorithm uses entropy to calculate the homogeneity of a sample**.

19. In a decision tree, define knowledge gain.

Ans:- The information gained in the decision tree can be defined as **the amount of information improved in the nodes before splitting them for making further decisions**.

20. Choose three advantages of the decision tree approach and write them down.

Ans:-   
**It can handle both continuous and categorical variables**. The execution of a Decision tree algorithm must be possible without having to scale the data as well. While utilizing the decision tree algorithm, it is not necessary to credit the missing values.

21. Make a list of three flaws in the decision tree process.

Ans:- One of the limitations of decision trees is that **they are largely unstable compared to other decision predictors**. A small change in the data can result in a major change in the structure of the decision tree, which can convey a different result from what users will get in a normal event.

22. Briefly describe the random forest model.

Ans:- The random forest is **a classification algorithm consisting of many decisions trees**. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.