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

Supervised learning, also known as supervised machine learning, is **a subcategory of machine learning and artificial intelligence**. It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately.

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

One of the chief ML applications in healthcare is the identification and diagnosis of diseases and ailments which are otherwise considered hard-to-diagnose.

**3. Give three supervised learning examples.**

Regression , Classification and **Naive Bayesian Model.**

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

**Classification** : It involves grouping the data into classes.  If you are thinking of extending credit to a person, you can use classification to determine whether or not a person would be a loan defaulter. When the supervised learning algorithm labels input data into two distinct classes, it is called binary classification. Multiple classifications means categorizing data into more than two classes.

**Regression** : in regression, In regression, a single output value is produced using training data. This value is a probabilistic interpretation, which is ascertained after considering the strength of correlation among the input variables. For example, regression can help predict the price of a house based on its locality, size, etc.

5. Give some popular classification algorithms as examples.

Logistic Regression, Naïve Bayes, Stochastic Gradient Descent, K-Nearest Neighbours

6. **Briefly describe the SVM model.**

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?**

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**.

Support vectors are **data points that are closer to the hyperplane and influence the position and orientation of the hyperplane**.

9. **In the SVM model, define the kernel.**

A kernel is **a function used in SVM for helping to solve problems**. They provide shortcuts to avoid complex calculations. The amazing thing about kernel is that we can go to higher dimensions and perform smooth calculations with the help of it. We can go up to an infinite number of dimensions using kernels.

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

The effectiveness of SVM depends on **the selection of kernel, kernel's parameters and soft margin parameter C**

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

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

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

**Long training time for large datasets**. Difficult to understand and interpret the final model, variable weights and individual impact. Since the final model is not so easy to see, we can not do small calibrations to the model hence its tough to incorporate our business logic

**13. Notes should be written on**

1. The kNN algorithm has a validation flaw : yes

2. In the kNN algorithm, the k value is chosen.

Yes, Data scientists usually choose as an odd number if the number of classes is 2 and another simple approach to select k is set k=sqrt(n).

3. A decision tree with inductive bias

**Shorter trees are preferred over longer trees**. Trees that place high information gain attributes close to the root are preferred over those that do not.

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

* **Simple** to implement and intuitive to understand
* 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.
* **No Training Time for classification/regression :**The KNN algorithm has no explicit training step and all the work happens during prediction
* **Constantly evolves with new data**: Since there is no explicit training step, as we keep adding new data to the dataset, the prediction is adjusted without having to retrain a new model.
* **Single Hyperparameters**: There is a single hyperparameter, the value of K. This makes hyper parameter tuning easy.
* **Choice of distance metric**: There are many distance metrics to chose from. Some popular distance metrics used are Euclidean, Manhattan, Minkowski, hamming distance eand so on.

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

* **High prediction complexity for large datasets:**Not great for large datasets**,** since the entire training data is processed for every prediction. **Time complexity** for each prediction is **O(MNlog(k))** where M is the dimension of the data, N is the size or the number of instances in the training data. Note that there are specialized ways of organizing data to address this issue and make KNN faster.
* **Higher prediction complexity with higher dimensions:** The prediction compleixty in supervised learning gets higher for higher dimensional data (see the dependence of time complexity from the previous point on the dimension d).
* **KNN Assumes equal importance to all features:** Since KNN expects points to be close in ALL dimensions, it might not consider points that are really close in sevaral dimensions, though farther away in a few favourably. This can be adjusted by chosing an appropriate distance measure. Moreover, this means it is sensitive if different features have different ranges. This can be addressed by appropriate pre-processing to scale features.
* **Sensitive to outliers:**A single mislabeled example can change the class boundaries. This could specially be a bigger problem for larger dimensions, if there is an outlier in one dimension, since the average separation tends to be higher for higher dimensions (curse of dimensionality), outliers can have a bigger impact.

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

A decision tree is **a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks**. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.

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

**A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy).** **Leaf node (e.g., Play) represents a classification or decision**.

18. **What is a decision tree's entropy?**

Entropy is **an information theory metric that measures the impurity or uncertainty in a group of observations**. It determines how a decision tree chooses to split data.

**19. In a decision tree, define knowledge gain.**

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.**

* Simple to Understand for Coders.
* Missing Values Aren't an Issue.
* They are Inexpensive.

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

* Predictions Are Not Smooth or Continuous.
* Not good for Regression.
* Expensive

**22. Briefly describe the random forest model.**

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