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

Supervised machine learning is a method of solving machine learning problems by training a machine learning model using a labelled data. Labelled data would mean that for every nth set of values for i number of independent variables, a value x would be present belonging to the dependent variable where, value of x can belong to countably finite or uncountably infinite set.

Significance of the name is that the name itself explains that there will be some form of supervision working on machine learning. The labelled data acts as a supervision for machine learning as class labels or target variable values have been manually fed or generated from past records.

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

Several surgeries and therapies like chemotherapy are last resort method to patients debilitating them. These surgeries often run a risk of death if a combination of parameters like oxygen saturation level are not within limits. Crucial time is saved by feeding the patient's readings against such parameters into supervised machine learning models that can classify the survivability of a patient post surgery fast.

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

1. Sentiment Analysis on a product's reviews to determine how well the product is doing on an e-commerce platform.
2. Spam detection to remove e-mails that are advertising based.
3. Prediction of salary

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

Classification is a task performed using supervised learning where the target or dependent variable values in the labelled data come from a countably finite set.

Regression is a task performed using supervised learning where the target or dependent variable values in the labelled data come from a countably infinite set.

**5. Give some popular classification algorithms as examples.**

1. Naive Bayes classifier
2. Random Forest classifier
3. Support Vector classifier.

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

SVM or Support Vector Machine model is a machine learning model that can be used for classification and regression model. It relies on finding a hyperplane and support vectors for classification such that the two data points on the outskirts of their own classes of linearly classifiable data. These support vectors act as margins on either side of the hyperplane. Support vectors are chosen such that the margin on either side of hyperplane is maximum.

Support Vector Machine models are not only used for binary linear classification, but can also be used for multiclass non linear classification. Support Vector algorithms use kernels like polynomial kernels, quadratic kernels, string kernels and genome kernels for kernelization which is just like feature transformation.

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

Zeta is the variable that becomes the cost of misclassification. Its value is determined by the distance of a misclassified point beyond or inside a margin. 0 value of zeta would mean correct classification while a non zero value would mean misclassification on one side of the main hyperplane.

**8. In the SVM model, define Support Vectors.**

In SVM model, the hyperplane is calculated such that it maximises the margin(margin is set such that one margin touches data point of one class and one margin touches data point of another class) or maximises the distance between hyperplane and margin and classifies the data points as widely as possible.

These margin calculations are done by setting margin only those data points that face the hyperplane and belong to different class. Such data points that can be set as margin to maximise the distance between hyperplane and margin and classifies the data points as widely as possible will be known as support vectors.

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

In the SVM model, different mathematical functions can be used in place of the (transpose(xi)\*x) of the dual problem for various implicit feature transformations. Such mathematical functions that can make SVM work with nearly any form of non linear classification and regression is called the kernel.

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

Dimensionality, value of C and kernel used are the 3 factors that influence SVM's effectiveness.

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

The benefits of using the SVM model are->

A. Works well with high dimensionality.

B. Outliers have little impact in case of kernel SVMs.

C. Implicit kernelization helps SVM work and come up with non linear decision surfaces.

D. High generalizability due to margins

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

The drawbacks of using the SVM model are->

A. Finding or coming up with problem specific kernels for best performance of SVM is difficult.

B. No inherent way of getting feature importance.

C. Built model has low interpretability

D. Presence of two hyperparameters makes tuning time large.

**13. Notes should be written on**

**1. The kNN algorithm has a validation flaw.**

K nearest neighbors (KNN) is a supervised machine learning algorithm. A supervised machine learning algorithm’s goal is to learn a function such that f(X) = Y where X is the input, and Y is the output. KNN can be used both for classification as well as regression.

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

k is a hyperparameter in the KNN algorithm where k is the number of clusters that will be made. Optimal k value can be found using the elbow method in which, the performance of the clustering algorithm is measured against the values of k.

**3. A decision tree with inductive bias**

The inductive bias in decision tree is that shorter trees are preferred over longer trees. Smaller trees would mean smaller depths with small number of nodes.

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

The benefits of the KNN algorithm are->

A. Built model is easy to interpret.

B. Time and Space complexity low.

C. Train time is virtually none.

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

Some of the KNN algorithm's drawbacks are->

A. Sensitive to outliers.

B. Does not work well with high dimensionality.

C. Sensitive to missing values.

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

Decision Tree algorithm is a supervised machine learning algorithm where an inverted tree is created and data is split into nodes based on a threshold or condition passed or failed.

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

Node in a decision tree is a collection of data points for which decision has to be made whether to split or not, and how to split using a mathematical function.

Leaf in a decision tree is a terminating node or a terminating vector, beyond which no splits will be done.

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

Entropy is the measure of a purity or homogenity of a sample of data points. Using entropy, the best split of sample can be calculated.

Maximum entropy is 1 and is calculated for sample of data that would have 50-50% class label data points. Minimum entropy is 0 and is calculated for sample of data that would have 0% data points for class label 1 but 100% data points for class label 2, meaning that it will not be split further.

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

Numerical value used to quantify the quality of a split is called knowledge gain or information gain. It is calculated by subtracting the weighted entropies for each branch from the original entropy. Maximising information gain leads to best results in decision trees.

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

Three advantages of the decision tree approach are->

A.A function in the form of gini index is used and solved for getting mathematically sound splits and end classification.

B. Decision Trees can be visually realized, and have high interpretability.

C. Decision Tree approach is not distance based, only based on order and logic. Due to this, many preprocessing methods need not be applied.

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

Three flaws in the decision tree process are->

A. Decision Trees are inherently overfitting.

B. They do not perform well for high dimensional data.

C. Training an accurate model for continous data is difficult as little changes in the training data would lead to different results in classification on test data.

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

Random forest model is an aggregation ensemble model technique based on decision trees. Multiple base learner decision trees are trained on subsets of data and each comes up with its own classification result. At the end, a majority vote is used to come to the final conclusion regarding the true classification result.