**1. What is the definition of a target function? In the sense of a real-life example, express the target function. How is a target function's fitness assessed?**

**Ans1:-** Formulation by a machine learning algorithm which has been created using training data to predict target values based on independent variable input is called target function. It can also be called the method used to solve a problem statement, that has been created by an algorithm after training on training data.

Ensemble models for credit card fraud detection produce a target function for multiple models after training phase. These target function are then used in real life to produce result or prediction based on the inputs they gets from the real world.

A target function's fitness can be assessed using cross validation and evaluation metrics.

**2. What are predictive models, and how do they work? What are descriptive types, and how do you use them? Examples of both types of models should be provided. Distinguish between these two forms of models.**

**Ans2:-** Predictive models are machine learning models that use training data to train themselves and predict target value based on the independent variable input given to them. Prediction can be in the form of classification and regression. Example: Prediction of future salary based on experience can be done using Linear Regression.

Descriptive models are built to identify trends and underlying patterns. Most of descriptive models are built using unsupervised machine learning. Groups can be created by clustering together individual data points using clustering machine learning models. When clustering is done, feature importance can be used to find the variables contributing most to the clustering and on the basis of that, a working theory or hypothesis can be made about underlying trends and patterns. Example: Clustering can be done using individuals from different countries to find underlying trend about differences in life expectancies and mortality rates.

**3. Describe the method of assessing a classification model's efficiency in detail. Describe the various measurement parameters.**

**Ans3:-** After training the machine learning model on training data, it can be tested against test data. Its performance on this unseen data is measured using appropriate performance metric. Different performance metrics exist because a single metric cannot be used for measurement in all cases. For example, accuracy score can give misleading results if used on models trained and tested on imbalanced data.

Recall score: Defined as number of true positives divided by total actual positive points, recall score is used to measure sensitivity of a model. It can also be defined as correctly detected true positives out of all positive class points. It can only be used on binary classifiers.

Precision score: Defined as number of true positives divided by summation of all true positives and false positives. In other words, of all the points model predicted positive how many were actually positive. It can only work on binary classifiers.

Fl score: Defined as 2 (pr re)/(pr+re) where pr is precision score and re is recall score. Works only on binary classifiers.

ROC\_AUC score: Receiver Operating Characteristics-Area Under Curve score is derived from calculating the area under the curve plotted between True Positive Rate and False Positive Rate. True Positive Rate is simply recall score and False Positive Rate is total false positives divided by total number of actual negatives. ROC\_AUC score ranges from 0 to 1 where 0 is bad and 1 is good. This score is impacted by models that are too simple. Extensions can be made to ROC\_AUC score to make it capable of assessing multi class classifiers.

Log Loss: Used for both binary classifier and multiclass classifier. Can only be used on models that are based on probability or produce probability scores to classify data points. Log loss should be as low as possible and close to 0. Range of Log loss is from 0 to +inf.

**4. i. In the sense of machine learning models, what is underfitting? What is the most common reason for underfitting?**

**Ans4:-** Underfitting is a situation arising when the hypothesis is way too simple, has poor generalizability, and when the machine learning model is way too simple to produce good results. Underfitting causes a model to produce poor results due to heavily simplified algorithm reacting lightly to changes in the unseen data for independent variables from the training data. Underfitting is also called High Bias.

Common reason for underfitting is presence of too many features in the dataset.

**ii. What does it mean to overfit? When is it going to happen?**

Overfitting is a situation arising when the hypothesis is way too complex, or when the machine learning model is way too complex to produce good results. Overfitting makes a model produce poor results due to slightest variations in the unseen data for independent variables from the training data. Overfitting is also called High variance.

Common reason for overfitting is small dataset.

**iii. In the sense of model fitting, explain the bias-variance trade-off.**

Bias means simplifying the model so that the resultant target function has better generalizability. Variance means change in result from target function when training data differs. If during model fitting, model is kept too simple and has few parameters then the resultant model is high bias and low variance model. If during model fitting, model is kept too complex and has large number of parameters then the resultant model is high variance and low bias model. Either of the two conditions leads to incorrect results after training phase. We need to balance between these two situations by using techniques that minimize both high variance and high bias. This is the bias-variance tradeoff.

**5. Is it possible to boost the efficiency of a learning model? If so, please clarify how.**

**Ans5:-**

Using ensemble models:--- Ensemble models use techniques like bagging and boosting to reduce bias and variance while improving the performance.

Parameter tuning:--- Parameters or model hyper-parameters must be set manually for optimal performance. Values for these hyper-parameters can be found using Brute Force techniques like Grid Search technique with ease.

Feature engineering:--- Feature engineering is a very important part of machine learning. Transforming datapoints into higher or lower dimension, or simply into features that can be used better for training phase and creating new features from existing features is called Feature Engineering. Support Vector Classifier uses kernels for implicit feature transformations, also called kernelization. Techniques like SMOTE is used for creating new datapoints.

Feature selection:--- Using feature importance, features can be selected based on the information they provide for better model building. Other techniques are forward feature selection and backward feature selection. In forward feature selection, features are added one by one and model is trained and tested each time to measure increase or decrease in performance. The exact opposite happens in backward feature selection.

**6. How would you rate an unsupervised learning model's success? What are the most common success indicators for an unsupervised learning model?**

**Ans:6-** Rating an unsupervised learning model's success can be done using several metrics that rely on calculating inter cluster, intra cluster distances, probabilistic measures etc.

Most common indicators for an unsupervised learning model are silhouette width, Dunn index and Davies Bouldin index.

**7. Is it possible to use a classification model for numerical data or a regression model for categorical data with a classification model? Explain your answer.**

**Ans7:-** Classification models can be used for numerical data but only in certain cases, with certain modifications. The target variable values can be put into bins. After these bins have been created and all values have been changed to bins in which the original values lie, a multiclass classification model can be trained and tested. This process is also called discretization.

For instance, if Sepal length is to be predicted using a classification model in IRIS dataset, we can first one hot encode the variety or species column. Then, values (4,4.2,4.5,4.7,5) can be grouped into one category labelled 0. This can go on until we exhaust all Sepal length training values.

Decision Trees can also handle both classification and regression in the form of CART algorithm.

Regression models, in even rarer cases can be used for classification purpose. Linear Regression, for example can be used by specifying a threshold value below and beyond which, data points are classified in a binary fashion.

**8. Describe the predictive modeling method for numerical values. What distinguishes it from categorical predictive modeling?**

**9. The following data were collected when using a classification model to predict the malignancy of a group of patients' tumors:**

**i. Accurate estimates – 15 cancerous, 75 benign**

**ii. Wrong predictions – 3 cancerous, 7 benign**

Determine the model's error rate, Kappa value, sensitivity, precision, and F-measure.

Ans9:

In\_accurate\_yes=3

In\_accurate\_no=7

Actual\_yes=22

Actual\_no=78

error\_rate=(In\_accurate\_yes+In\_accurate\_no)/(Actual\_yes+Actual\_no)

print(error\_rate) = 0.1

agreed=(15+75)/100

p\_yes=((15+7)/100)\*((15+75)/100)

p\_no=((3+75)/100)\*((3+7)/100)

print(agreed, p\_yes, p\_no) = 0.9 0.198 0.07800000000000001

non\_dis = p\_yes+p\_no

print(1-agreed)

print(1-non\_dis)

0.09999999999999998

0.724

k=1-(0.09999999999999998/0.724)

print(k) #Kappa value

0.861878453038674

re=(15/(15+7)) # Sensitivity Score

print(re)

0.6818181818181818

pr=(15/(15+3)) #Precision Score

print(pr)

0.8333333333333334

print(2\*((pr\*re)/(pr+re))) # F1 score

0.7499999999999999

10. Make quick notes on:

Ans10:-

1. The process of holding out

Hold out process is simply holding a small portion of data as unseen data using which the trained model will be tested.

2. Cross-validation by tenfold

Ten-fold Cross-validation is a technique used to check the efficiency of machine learning models by sub setting data into 10 equal subsets. One subset is set aside for validation/testing and 9 subsets are used as training data. Cross validation is then done 10 times with each subset of the training data and tested against test data. Average of the all test scores is taken as the final test score for the model.

3. Adjusting the parameters

Model hyper-parameters or simply parameters are values that have to be manually specified for best performance. Values for these parameters is found by employing Brute Force techniques. At the chosen value, Cross validation error would be very low. As hyper-parameters increase, time and space complexity increases exponentially.

11. Define the following terms:

Ans11:-

i. Purity vs. Silhouette width

Purity is an external metric for measuring performance of clusters. It can be defined as the number of data points that were classified correctly into clusters.

Estimate of average inter cluster distance to give efficacy/performance of cluster algorithms is called width of the silhouette. Its value ranges from -1 to 1 where 1 means good and -1 means bad.

ii. Boosting vs. Bagging

Boosting is one of the ensemble machine learning modelling techniques in which, multiple base learners are trained in series over training data. Base learners are added to overcome the errors in previously added and trained base learners, until the entire training data has been correctly predicted. It reduces bias.

Bagging is one of the ensemble machine learning modelling techniques in which, multiple base learners are trained in parallel over subsets of training data, sampled with replacement. All base learners produce a prediction and after a majority vote over the prediction, final value is output as the result. It reduces variance without impacting bias.

iii. The eager learner vs. the lazy learner Eager learner algorithms are ones that generate a generalizable target function during the training phase. Unlike lazy learners where the computation over data is deferred until a new instance or query is input after which comparison between stored examples and query takes place, eager learners use the target function generated during training phase to generate the result for a new query. Example: Logistic Regression.

Lazy learner or lazy learning algorithms, also called instance based learning algorithms are the ones which use stored instances seen during training phase of the model and then compare new instances with the stored instances to produce a result or prediction. These algorithms are called lazy learners because the actual operations or processing over data happens only when the trained model gets new instance for which a result has to be output. Example: K-Nearest Neighbor algorithm.