1. Give short description each of Linear, RBF, Polynomial kernels used in SVM.

Ans: **Linear Kernel** is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are a Large number of Features in a particular Data Set. One of the examples where there are a lot of features, is Text Classification, as each alphabet is a new feature. So we mostly use Linear Kernel in Text Classification.

**RBF kernel,** (Radial Basis Function Kernel). Suppose we have three classes they will not be linearly separable, So, what this kernel basically does is that it tries to transform the given data into almost linearly separable data. Radial Basis Function Kernel (RBF): The similarity between two points in the transformed feature space is an exponentially decaying function of the distance between the vectors and the original input space as shown below. RBF is the default kernel used in SVM.

**Polynomial Kernel:** The Polynomial kernel takes an additional parameter, ‘degree’ that controls the model’s complexity and computational cost of the transformation.

1. R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit of model in regression and why??

Ans: R-squared indicates how well a regression model fits a data set. However, it doesn’t tell us the entire story. To get the full picture, we must must consider R2 values in combination with residual plots, as the residual sum of squares measures the amount of error remaining between the regression function and the data set. A smaller residual sum of squares figure represents a regression function. Residual sum of squares–also known as the sum of squared residuals–essentially determines how well a regression model explains or represents the data in the model.

1. What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other?

Ans: The Total SS (TSS or SST) tells you how much variation there is in the dependent variable.

Total SS = Σ(Yi – mean of Y)2

**Explained sum of square (ESS)** is a statistical quantity used in modelling of a process. ESS gives an estimate of how well a model explains the observed data for the process.

**ESS = total sum of squares – residual sum of squares**

A **residual sum of squares** (**RSS**) is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model

RSS=∑ni=0(ϵi)2=∑ni=0(yi−(α+βxi))2

1. **What is Gini –impurity index?**

Ans: Gini Impurity is a measurement of the likelihood of an incorrect classification of a new instance of a random variable, if that new instance were randomly classified according to the distribution of class labels from the data set.

1. Are unregularized decision-trees prone to overfitting? If yes, why?

Ans: They are prone to this because they are very data intensive - that is, they examine the data in a lot of ways. At each node, they look at every possible split of every independent variable, Even with a relatively small number of variables, that can be a lot of things to examine, especially if one of them is a categorical variable with more than a few levels.

1. What is an ensemble technique in machine learning?

Ans: Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Ensemble methods usually produces more accurate solutions than a single model would.

1. What is the difference between Bagging and Boosting techniques?

Ans: Bagging is a way to decrease the variance in the prediction by generating additional data for training from dataset using combinations with repetitions to produce multi-sets of the original data. Boosting is an iterative technique which adjusts the weight of an observation based on the last classification.

**Advantages of Bagging:**

Reduces over-fitting of the model.

Handles higher dimensionality data very well.

Maintains accuracy for missing data.

**Advantages of Boosting:**

* Supports different loss function
* Works well with interactions.

1. what is out-of-bag error in random forests?

Ans: Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging) to sub-sample data samples used for training.

1. What is K-fold cross-validation?

Ans: Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation. It is a popular method because it is simple to understand and because it generally results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.

The general procedure is as follows:

Shuffle the dataset randomly.

Split the dataset into k groups

For each unique group:

Take the group as a hold out or test data set

Take the remaining groups as a training data set

Fit a model on the training set and evaluate it on the test set

Retain the evaluation score and discard the model

Summarize the skill of the model using the sample of model evaluation scores

1. What is hyper parameter tuning in machine learning and why it is done?

Ans: **Hyperparameter optimization** or **tuning** is the problem of choosing a set of optimal **hyperparameters** for a **learning** algorithm. A **hyperparameter** is a **parameter** whose value is used to control the **learning** process.

**Hyperparameters** are **important** because they directly control the behaviour of the training algorithm and have a **significant** impact on the performance of the model is being trained.

1. What issues can occur if we have a large learning rate in Gradient Descent?

Ans: When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error.

1. What is bias-variance trade off in machine learning?

Ans: Bias is the simplifying assumptions made by the model to make the target function easier to approximate. Variance is the amount that the estimate of the target function will change given different training data. Trade-off is tension between the error introduced by the bias and the variance.

1. What is the need of regularization in machine learning?

Ans: Regularisation is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

The commonly used regularisation techniques are :

L1 regularisation

L2 regularisation

Dropout regularisation

1. Differentiate between Adaboost and Gradient Boosting?

Ans: In Adaboost, ‘shortcomings’ are identified by high-weight data points. In Gradient Boosting, ‘shortcomings’ (of existing weak learners) are identified by gradients.

Gradinet Boosting is an approach trains learner based upon minimising the loss function of a learner. Adaptive Boosting method focuses on training upon misclassified observations. Alters the distribution of the training dataset to increase weights on sample observations that are difficult to classify.

1. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Ans: Logistic regression has traditionally been used to come up with a hyperplane that separates the feature space into classes. But if we suspect that the decision boundary is nonlinear we may not get a better results by attempting some linear functional, instead if we use some non-linear functional forms for the logit function.