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

Ans : In machine learning, the radial basis function **kernel**, or **RBF kernel**, is a popular **kernel** function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification.

In machine learning, the **polynomial kernel** is a **kernel** function commonly used with **support vector machines** (**SVMs**) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over **polynomials** of the original variables, allowing learning of non-linear models.

**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. ... Training a **SVM** with a **Linear Kernel** is Faster than with any other **Kernel**

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??

A residual sum of squares (RSS) is a statistical technique used to measure the amount of [variance](https://www.investopedia.com/terms/v/variance.asp) in a data set that is not explained by a regression model. Regression is a measurement that helps determine the strength of the relationship between a dependent variable and a series of other changing variables or independent variables.

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.

The residual sum of squares is used to help you decide if a statistical model is a good fit for your data. It measures the overall difference between your data and the values predicted by your estimation model (a “residual” is a measure of the distance from a data point to a regression line). Total SS is related to the total sum and explained sum with the following formula:

Total SS = Explained SS + Residual Sum of Squares.

1. What is Gini –impurity index?

The **Gini impurity** measure is one of the methods used in decision tree algorithms to decide the optimal split from a root node, and subsequent splits. ... Def: **Gini Impurity** tells us what is the probability of misclassifying an observation. Note that the lower the **Gini** the better the split

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

6.What is an ensemble technique in machine learning?

**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. ... These models, when used as inputs of **ensemble methods**, are called ”base models”.

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

**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.

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

**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.

9. What is K-fold cross-validation?

**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.

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

In **machine learning**, **hyperparameter optimization** or **tuning** is the problem of choosing a set of optimal **hyperparameters** for a **learning** algorithm. ... These measures are called **hyperparameters**, and have to be tuned so that the model can optimally solve the **machine learning** problem.

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

A **learning rate** that is too small may never converge or may **get** stuck on a suboptimal solution. **When** the **learning rate** is too **large**, **gradient descent can** inadvertently increase rather than decrease the **training** error

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

In [statistics](https://en.wikipedia.org/wiki/Statistics) and [machine learning](https://en.wikipedia.org/wiki/Machine_learning), the **bias–variance tradeoff** is the property of a set of predictive models whereby models with a lower [bias](https://en.wikipedia.org/wiki/Bias_of_an_estimator) in [parameter](https://en.wikipedia.org/wiki/Statistical_parameter) [estimation](https://en.wikipedia.org/wiki/Estimation_theory) have a higher [variance](https://en.wikipedia.org/wiki/Variance) of the parameter estimates across [samples](https://en.wikipedia.org/wiki/Sample_(statistics)), and vice versa. The **bias–variance dilemma** or **bias–variance problem** is the conflict in trying to simultaneously minimize these two sources of [error](https://en.wikipedia.org/wiki/Errors_and_residuals_in_statistics) that prevent [supervised learning](https://en.wikipedia.org/wiki/Supervised_learning) algorithms from generalizing beyond their [training set](https://en.wikipedia.org/wiki/Training_set)[[1]](https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff#cite_note-1)[[2]](https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff#cite_note-2):

* The [*bias error*](https://en.wikipedia.org/wiki/Bias_of_an_estimator) is an error from erroneous assumptions in the learning [algorithm](https://en.wikipedia.org/wiki/Algorithm). High bias can cause an algorithm to miss the relevant relations between features and target outputs (underfitting).
* The [*variance*](https://en.wikipedia.org/wiki/Variance) is an error from sensitivity to small fluctuations in the training set. High variance can cause an algorithm to model the random [noise](https://en.wikipedia.org/wiki/Noise_(signal_processing)) in the training data, rather than the intended outputs ([overfitting](https://en.wikipedia.org/wiki/Overfitting" \o "Overfitting)).

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

**Regularization** is a technique used for tuning the function by adding an additional penalty term in the error function. The additional term controls the excessively fluctuating function such that the coefficients don't take extreme values

14. Differentiate between Adaboost and Gradient Boosting

**Adaboost** is more about 'voting weights' and **gradient boosting** is more about 'adding **gradient** optimization'. **Adaboost** doesn't overfit because it is more about 'organizing people to vote' than 'voting'. In fact, if you have a **gradient boosting** model, you can use it in **adaboost** along with other models.

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

**Logistic regression is** known and used as a **linear classifier**. It **is** used to come up with a hyperplane in feature space to separate observations that belong to a class from all the other observations that do **not** belong to that class