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

Ans:

R2:

R2 provides a useful measure of how well a model fits, in terms of (squared) distance from points to the best fitting line. However, as one adds more regression coefficients, R2 never goes down, even if the additional X variable is not useful. In other words, there is no adjustment for the number of parameters in the model.

Adjusted R2:

In a simple linear regression, where p, the number of independent variables is one, then Adjusted R2 = R2 . As the number of parameters increases, Adjusted R2 ≤ R2 , with this definition: R 2 = 1 − (n − 1) × sum of squared residuals from model with α and β (n − p) × sum of squared residuals from model with α only So, there is some attempt to adjust for the number of parameters.

2.**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: TSS(Total Sum of Squares) gives us how much variation in the observed data. RSS (Residual Sum of Squares)measures the variation in the error between the observed data and ESS(Explained Sum of squares) gives us the value estimated by the regression

 Equation: TSS=ESS+RSS.

3.**What is the need of regularization in machine learning?**

Ans: To Make sure our models doesn't overfit and to reduce the erros,regularization technique is used in Machine Learning.

4. **What is Gini–impurity index?**

Gini Index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.

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

 Ans: Decision Tree is prone to overfit model,when the tree is particularly Deep.so The Chance of Error may be high ,It takes time and this can be reduced only by regularizing the model by adjusting the Parameter like Max depth. and choose correct way of decision like GINI or Entropy methods leaf size .Hence regularisation is very important in Decision Tree ,otherwise model overfit.

**6. What is an ensemble technique in machine learning?**

Ans:

The Decision taken should not be based on the Maximum features, for example there are so many TV shows where the final decision of the Program will be decided by the Poll given by people, where it would not be bizard and can taken from the Maximum votes likes. In the same way ensemble Technique acts like .and we have 2 categories like Bagging and Boosting.

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

Ans: Suppose we have 5 models like giving decision1.yes,2.No,3.Yes,4.No,5.Yes

Bagging is a technique which is used to work on all 5 models Parallel and gives the final decision As the most getting one.

Boosting is a technique that works sequentially on the given models and gives one as Final answer.

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

Ans: The *out-of-bag* (OOB) error is the average error for each  calculated using predictions from the trees that do not contain in their respective bootstrap sample. This allows the `Randomclassifier` to be fit and validated whilst being trained

**9. What is K-fold cross-validation?**

Ans:In the given Model,we choose cv=5 as parameter,

K fold cross validation will train the model 5 times ,split in to 5 pieces.

It will choose (k-1) which is 4 here for training Datasets.and 1 dataset for testing Purpose.I works completely on 5 different data

It is costly method of cross validation and it takes lot of time to do the same,

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

*Hyperparameters*, that cannot be directly learned from the regular training process. They are usually fixed before the actual training process begins. These parameters express important properties of the model such as its complexity or how fast it should learn.
Some examples of model hyperparameters include:

1. The penalty in Logistic Regression Classifier i.e. L1 or L2 regularization
2. The learning rate for training a neural network.
3. The C and sigma hyperparameters for support vector machine.
4. The k in k-nearest neighbors.

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

A learning rate that is too large can **cause the model to converge too quickly to a suboptimal solution**, whereas a learning rate that is too small can cause the process to get stuck.

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

logistic Regression has traditionally been used as a linear classifier, i.e. when the classes can be separated in the feature space by linear boundaries. That can be remedied however if we happen to have a better idea as to the shape of the decision boundary.

## 13. Differentiate between Adaboost and Gradient Boosting.

*Ada Boost*                                    *Gradient Boosting.*

| Ada Boost | Gradient Boosting |
|---|---|
| 1 Minimises the exponential loss function that makes algorithm most sensitive to the outliers. | More robust to the outliers. |
| 2.Its first designed boosting algorithm | It is more flexible. |
| 3.Designed for Binary classification problems. | It is used for both classification and regression problems. |

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

In statistics and machine learning, the bias–variance tradeoff is **the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters**.

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

LINEAR KERNELS:

Linear SVM is used for **linearly separable data**, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

RBF KERNELS:

: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.

POLYNOMIAL KERNELS:

A polynomial kernel is **a kind of SVM kernel that uses a polynomial function to map the data into a higher-dimensional space**. It does this by taking the dot product of the data points in the original space and the polynomial function in the new space.