Q1. 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 1: RSquare is a metric which is used mostly in simple regression problem to check how our model is good. RSquare is better measure of goodness of fit the model because in Rsqaure we found RSS and TSS and on the basis of these two values we check how good our model is.

RSS(Residual Sum of Square) : It found total number of error or residual (how much difference b/w actual and predicted value) square them and add them.

TSS(Total Sum of Square) : It found the average of actual values and subtract them from the actual values then square them and add them.

Formula : R2 = 1 – RSS / TSS

R2 values come b/w 0 – 1, if R2 metric result is close to 1 then your model is performing well.

Q2. 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 2 : **RSS(Residual Sum of Square)** : It found total number of error or residual (how much difference b/w actual feature values and predicted value) square them and add them.

TSS(Total Sum of Square) : it found the average distance b/w each actual value and predicted value then square them add them.

Formula :

ESS : It found the distance b/w the residual value and average value then square them and add them .

Equation : R2 = ESS/TSS

TSS=ESS+RSS

TSS/TSS=ESS/TSS + RSS/TSS

R2=ESS/TSS = 1 -RSS/TSS

Note : We use square in the equation because some time we got negative values so it

Convert into positive values.

Q3. What is the need of regularization in machine learning?

Ans 3 : Regularization is a process to avoid overfitting situation. In this we use few models Like Lasso and Ridge.

Q4. What is Gini–impurity index?

Ans : 4. It prioritized the feature from lower to higher on the basis of impurity, means which feature have less impurity (less waste information) about the label that will be the best feature.

Example : Suppose f1,f2,f3 are three features which have relation with label where f2 have 5% impurity,f1 have 12% impurity and f3 have 7% impurity. Then f2 will be the best feature because it has low impurity than another feature, then second lowest f3 and third lowest f1 and so on.

Q5. Are unregularized decision-trees prone to overfitting? If yes, why?

Ans: Yes, Decision trees are prone to overfitting when they are deep. So, in this situation we regularized the Decision tree by giving Maximum Depth of the tree.

Q6. What is an ensemble technique in machine learning?

Ans: These techniques combined multiple models and make the decision. It is two types.
1. Bagging
2. Boosting

Q7. What is the difference between Bagging and Boosting techniques?

Ans: **Bagging:** In bagging, model work parallelly means all models work together at the same time. it will take the decision on the basis of model's majority in classification problem and in regression problem it gives their decision on the model's average value

Boosting : In boosting, models work sequentially means one model work at a time and give their output that output will be share by the second model it will give their output and at the end it will give their decision on the basis of which category or class all the model have higher percentage not on the basis of Majority.

Q8. What is out-of-bag error in random forests?

Ans: In Bagging technique we use multiple models to take the decision so we share feature data to the model but some portion of data or sample of feature are not given to the models so that portion of data is called out-of-Bag-Evaluation which is used to evaluate the model or testing the model.

Q-9. What is K-fold cross-validation

Ans: First, Cross Validation is use to increase the model accuracy or also check our model is overfitted or not.

Now, K-fold cross validation, it makes suppose 5 copy or 5 iterations where all iterations have 100% data of dataset and in each iteration 100% data will be divided into 5 parts where 80% is use for training and 20% for testing and now each iteration will give their result where we find the mean value of all iteration. Then check is mean value near to the model accuracy value. if it is near means model is not overfitted if not then model can be overfitted.

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

Ans: When we tune the parameter or give user define values to the parameter of the model is called Hyper Parameter Tunning.

Model used default parameter or default values to build the model but when we use tunned parameters of model rather than default parameter model performance can be better.

It is used to avoid Biasness and High variances problems so that our model cannot be underfitted and overfitted.

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

Ans: If learning rate is high Model may not be reach to their target or not understand the data properly or may be overshotted.

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

Ans: No, because It is a binary classification algorithm which is used only for definite class or category and Regression problem work with the continuous data.

Q13. Differentiate between Adaboost and Gradient Boosting

Ans: **AdaBoost:**

1. An additive model where shortcomings of previous models are identified by high-weight data points.
2. The trees are usually grown as decision stumps.
3. Each classifier has different weights assigned to the final prediction based on its performance.
4. It gives weights to both classifiers and observations thus capturing maximum variance within data.

 Gradient Boost:

1. An additive model where shortcomings of previous models are identified by the gradient.
2. The trees are grown to a greater depth usually ranging from 8 to 32 terminal nodes.
3. All classifiers are weighed equally and their predictive capacity is restricted with learning rate to increase accuracy.
4. It builds trees on previous classifier's residuals thus capturing variance in data.

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

Ans: Bias: When Model give more importance or very simplistic about only few feature rather than other feature is also important then it is called biasness and if our model is too biased then model can be underfitted.

Variance: Variances leads to overfitting, means when there is high variance between the features data, Model can be overfitted.

Q-15: Give short description each of Linear, RBF, Polynomial kernels used in SVM

Ans: **Linear SVM**: Linear SVM is used when the data is divided in such a way it separated linearly. Means Datapoint can be classified into two classes by using a straight line.

We know what is kernel: It is used in non-linear data, what will do when data is not linearly separated. then we use Kernal trick where it will square the higher category datapoint of a class to increase the number of datapoint so that it can be go upper where other class datapoint will remain same. Now we can draw a line that is Hyperplane.

Polynomial and RBF Kernal was not taught us yet so we have to learn first then We can answer it.