MACHINE LEARNING

Q1 to Q15 are subjective answer type questions, Answer them briefly.

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?
   Residual Sum of Squares (RSS) is better measure of goodness of fit as R-squared value increases whenever a new feature is taken for consideration even if the feature is unrelated to the response. However, RSS controls this overfitting.

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.
   Total Sum of Squares (TSS) - It determines the total amount of variability in the dependent variable i.e., it measures how much the data points varies from its mean. A higher sum of squares indicates higher variability while a lower result indicates low variability from the mean.
   Equation- $\sum(Yi - \bar{Y})^2$

   Explained Sum of Squares (ESS) - It gives an estimate about how well a model represents the modeled data. It is also known as Regression sum of squares or Model sum of squares. A higher ESS indicates that the model does not do a good job fitting the data. A lower ESS indicates that the model does a good job fitting the data.
   Equation- $\sum(\hat{Y}i - \bar{Y})^2$

   Residual Sum of Squares (RSS) - It determines how much of the dependent variable's variation your model does not explain. It is the sum of the squared differences between the actual Y and the predicted Y. The smaller the residual sum of squares, the better your model fits your data and vice-versa.
   Equation- $\sum(Yi - \hat{Y}i)^2$

3. What is the need of regularization in machine learning?
   Regularization is a technique used to prevent overfitting of model by adding extra information to it. Sometimes the ML model performs well with the training data but does not perform well with the test data, i.e., it is not able to predict the output when it deals with unseen data, and hence the model is called overfitted. This problem can be dealt with the help of regularization technique. There are mainly two types of regularization techniques: Ridge and Lasso.

4. What is Gini–impurity index?
   Gini-impurity index is a measure of the impurity or disorder of a set of data in decision tee algorithms. It is used to determine the splitting criteria for creating branches in the tree.

5. Are unregularized decision-trees prone to overfitting? If yes, why?
   Yes, unregularized decision-trees are prone to overfitting because they have tendency to create complex trees that perfectly fit the training data but may not fit the test data or unseen data.

6. What is an ensemble technique in machine learning?
   Ensemble technique in machine learning involves combining multiple models to make accurate predictions. It helps to reduce bias and variance and increase overall stability of the model. Some popular ensemble methods include bagging, boosting and stacking.

7. What is the difference between Bagging and Boosting techniques?
   Bagging or bootstrap aggregating, involves training multiple models independently on different subsets of training data. Each model is trained on a randomly sampled subset of data with replacement. The predictions of these models are then combined by taking average or majority vote to make final prediction. It helps to reduce variance and improve model stability.
   On the other hand, boosting focuses on improving the performance of a single model. Models are trained in a sequence with each subsequent model trying to rectify the mistakes made by previous model. The models are weighted based on their performance, the better the performance the more the weight. This model helps to reduce bias and improve overall accuracy of the model.

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

   Out-of-bag error in random forests is an estimate of model's performance on unseen data. It is calculated by evaluating model's predictions on the training instances that were not included in bootstrap sample used to train each individual tree in the forest.

9. What is K-fold cross-validation?

   It is a technique used to assess the performance of a machine learning model. It involves splitting the data into k equal sized folds. The model is then trained and evaluated k times, with each fold being used as a validation set once while the remaining fold are used for training.

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

    Hyperparameter tuning refer to the process of finding the optimal values for the hyperparameters of a model. Hyperparameters are parameters that are not learned from data during training but are set before the training process begins. E.g., learning rate, no. of hidden layers in a neural network, etc.
    Hyperparameter tuning is done to improve the performance of machine learning model by finding the best combination of hyperparameters, enhancing model's ability to generalize and make accurate predictions on unseen data.

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

    The following issue may occur in case of a large learning rate in Gradient Descent:
    1. *Algorithm may fail to converge or take longer to converge.*
    2. *Algorithm become more sensitive to the noise in the data, causing it to converge to suboptimal solutions or get stuck in local minima.*

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

    No, Logistic Regression is not suitable for classification of non-linear data as it assumes linear relationship between the features and may not be able to capture the complex patterns to make accurate predictions.

13. Differentiate between Adaboost and Gradient Boosting.

    Adaboost or Adaptive boosting, focuses on adjusting the weights of training instances in each iteration to prioritize the misclassified instances. It trains weak learners sequentially with each subsequent learner placing more emphasis on the misclassified instances from previous iterations. It is effective in handling binary classification problems.
    On the other hand, gradient boosting builds the ensemble model by training weak learners in as stage-wise manner. Unlike Adaboost, gradient boosting does not adjust the weights of the training instances, instead it trains the weak learners to minimize the errors of previous weak learners. This is done by fitting the weak learners to the residuals (difference between actual and predicted values) of the previous learners.

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

    Bias-variance refers to the balance between the bias and variance of a model. Bias represents the error introduced by the model assumptions and simplifications. A high bias model tends to underfit the data and has a limited ability to capture the underlying patterns. It may oversimplify the relationship between the features and the target variable, resulting in poor performance on both the training and test data.
    Variance, on the other hand, represents the model's sensitivity to the fluctuations in the training data. A high variance model tends to overfit the data and performs well on the training data but poorly on the test data. It may capture the noise and randomness in the training data, leading to poor generalization to unseen data.
    The goal is to find the right balance between bias and variance. A model with too much bias will have limited predictive power, while a model with too much variance will be overly sensitive to the training data and fail to generalize as well.

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

    Linear kernel: The linear kernel is the simplest kernel used in Support Vector Machines (SVM). It assumes a linear relationship between the features and the target variable. It works well when the data is linearly separable, meaning the classes can be separated by a straight line or hyperplane.
    RBF (Radial Basis Function): It is capable of capturing non-linear relationships between the features and the target variable. It can handle complex decision boundaries and is effective for a wide range of applications.

Polynomial Kernels: The polynomial kernel is another non-linear kernel used in SVM. It maps the data into a higher-dimensional space using polynomial functions. The polynomial kernel is useful when the data exhibits non-linear patterns. The degree of the polynomial determines the complexity of the decision boundary.