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?

   Ans. Compared to Residual sum of squares and R-squared, R-squared measure is better measure of goodness of fit in regression because it represents the proportion of variance in the dependent variable which is explained by the independent variables. And on other hand it measures the square of difference between the observed data and predicted data which describes how well the model fits the data hence R-squared measure is a good choice for goodness of fit model in regression.
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. Total sum squares which tells about the total variability of dependent variables it is calculated by the sum of the squared difference between each independent variable value and the mean of the dependent variable, and it's is importance is decomposing the variance and understanding the proportion of variability explained by the regression model.

   ESS known as explained sum of squares also checks the variability of dependent variables in the regression explained by the regression model. The key role of the ESS is it tells how well the model fits the data by calculating the sum of squared differences between the predicted values and the mean of the dependent variable.

   RSS known as residual sum of squares and is a measure used in regression analysis to quantify the difference between the observed value of the dependent variable and the values predicted by the regression mode by calculating the sum of squared differences between the actual values and the predicted values. Lower the RSS value indicates the better fit.

   The equation relating the Three metrics with each other is:
   Total sum squares(TSS)= Explained sum of squares(ESS)+Residual Sum Of Squares(RSS).

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

   Ans. Regularization in machine learning is crucial because it prevents overfitting where the model performs well for the trained data and poorly on unseen data, with the help of L1 and L2 regularization by penalizing over complex improves the model performance. In the presence of highly correlated features it handle's multicollinearity issues. Overall it promotes simpler and more generalizable models, enhancing the models performance for unseen data.

4. What is Gini–impurity index?

   Gini-impurity index is a measure of how often a randomly selected element from the set would be incorrect labeled if it were randomly labeled according to the distribution of labels in the subset. It is calculated by summing the probability of each item being chosen times the probability of a mistake in categorizing that item.
5. Are unregularized decision-trees prone to overfitting? If yes, why?
   Ans. Yes unregularized decision-trees prone to overfitting. They tend to create complex, deep tress structures that make us to capture the noise or the outliers. for this the regularization techniques such as pruning or limiting tree depth are applied to prevent overfitting in the decision tress.
6. What is an ensemble technique in machine learning?
   Ans. To produce a more reliable and accurate model than any one model alone, ensemble techniques in machine learning combine the predictions of several models. Using model variety to increase overall performance is the fundamental notion behind ensemble approaches. Here are a few typical group methods.
   The Bootstrap Aggregating Bagging method: Idea: Use various subsets of the training data to train several instances of the same model. In order to aggregate the predictions of several decision trees trained on various subsets of the training data, Random Forest, for instance, uses a voting process.

Boosting: Concept: Train a number of inexperienced students in a sequential manner, paying attention to the errors committed by the earlier students. As an illustration, consider XGBoost, Gradient Boosting Machines (GBM), and Adaptive Boosting.

7. What is the difference between Bagging and Boosting techniques?
   Ans. These are the different ensemble technique that use multiple models to reduce the error and optimize the model. The bagging technique combines multiple models trained on the different subsets of data whereas the boosting trains the model sequentially , focusing on the error made by the previous model.
8. What is out-of-bag error in random forests?
   Ans. Out-of-bag error is a measure of ensemble used in random forest algorithms. It estimates the prediction error of the model without need of model without the need for a separate validation set.
9. What is K-fold cross-validation?
   Ans. In machine learning, K-fold cross-validation is a technique used to evaluate a model's performance and generalization. The model is trained and assessed k times, with the dataset being split up into "k" subsets. One of the subsets is utilized as the test set and the other k-1 subsets are used for training in each cycle. This procedure lessens the influence of randomness in a single split by ensuring resilience in evaluating a model's performance across several subsets of the data.

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

Ans. In machine learning, hyperparameter tuning is determining which set of hyperparameters will maximize a model's performance. Hyperparameters, which include learning rate, regularization strength, and the number of hidden layers in a neural network, are configuration options that are not learned from the input but are instead external to the model.
These hyperparameters are adjusted in order to improve the performance of a model. To determine the ideal configuration that optimizes the model's accuracy or minimizes its error on a specific job, many combinations are explored through iterative testing. This procedure helps the model become more effective on unknown data by enhancing its generalization capabilities.

11. What issues can occur if we have a large learning rate in Gradient Descent?
    Ans. When employing gradient descent with a high learning rate, it is possible to overshoot the cost function's minimum, which will cause the algorithm to diverge rather than converge. This could lead to oscillations, instability, or the inability to determine the model's ideal parameters. Selecting a suitable learning rate is essential to guarantee effective and steady convergence during training.

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
    Ans. Logistic regression is designed for linear classification problems, making it less suitable for handling non-linear data. When the relationship between features and the target variable is non-linear, logistic regression may struggle to capture complex patterns effectively. In such cases, more advanced techniques like decision trees, support vector machines, or neural networks are often preferred for handling non-linearities in the data.
13. Differentiate between Adaboost and Gradient Boosting.
    Ans. Both Adaboost and gradient boosting use a base weak learnerand they try to boost the performance of a weak learner by iteratively shifting the focus towards problematic observation that were difficult to predict. In adaboost shortcomings are identified by the high weighted data points whereas in gradient boosting shortcomings are identified by the gradients. During each iteration in adaboost the weights of the incorrectly classified samples are increased, so that the next weak learner focuses more on these samples whereas the gradient boosting updates the weight by computing the negative gradient of the loss function with respect to predicted output.
14. What is bias-variance trade off in machine learning?
    Ans. A crucial idea in machine learning that has to do with a model's performance is the bias-variance trade-off.When a complicated real-world problem is approximated using a simpler model, bias is the error that results. Excessive bias may cause underfitting, in which the model is unable to identify the fundamental patterns in the data.The sensitivity of the model to changes in the training set of data is known as variance. Excessive variance can lead to overfitting, a condition in which the model fits the training set too closely and performs poorly when applied to fresh, untested data. The trade-off arises from the fact that, increasing model complexity reduces bias but increases variance, and vice versa. Building models that perform effectively when generalized to new data requires finding the ideal balance. The model selected, its hyperparameters, and the quantity and caliber of training data all have an impact on this equilibrium.

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

Ans. Linear describes the kind of decision boundary that the algorithm seeks to establish between various classes in the data in Support Vector Machines (SVM). In order to efficiently divide data points from several classes, a linear SVM searches the feature space for a hyperplane. The main concept is to identify the best hyperplane that maximizes the margin—that is, the distance between each class's nearest data points, or support vectors—by measuring the hyperplane itself. When the dataset's classes are clearly divided in the feature space, this linear separation works well.

Support Vector Machines (SVM), "RBF" refers to "Radial Basis Function." For non-linear classification and regression problems in Support Vector Machines (SVMs), the Radial Basis Function is a widely utilized kernel function.

The radial basis function kernal is defined has the:

$$K(x,x') = \exp\left(-\frac{||x-x'||^2}{2\sigma^2}\right)$$

Here x and x' are the inputs and ||x-x'|| is the Euclidean distance and the σ width of the kernal.

"Polynomial" describes a kind of kernel function used in non-linear classification and regression applications within the context of Support Vector Machines (SVM). One of the kernel functions that enables SVMs to function in a higher-dimensional space and identify non-linear correlations in the data is the polynomial kernel.

The polynomial kernal is defined as

$$K(x,x') = (x^T x' + c)^d$$

Here the x and x' are the inputs and the c is constant representing the degree of the polynomial and the term in the bracket represents the dot product raised to d.