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 ofgoodness of fit model in regression and why?

   R-squared is a better measure of goodness of fit in regression. R-squared is a statistical measure that represents the proportion of variation in the dependent variable that is explained by the independent variables in the model. It is expressed as a percentage between 0% and 100%. The closer R-squared is to 100%, the better the model fits the data.On the other hand, the Residual Sum of Squares (RSS) measures the sum of the squared differences between the observed and predicted values. It is an absolute measure of the error in the model and does not provide any information about how well the model fits the data relative to other models. Additionally, the RSS value can be influenced by the scale of the dependent variable, making it less useful for comparing models.In summary, R-squared is a more intuitive and interpretable measure of goodness of fit, as it provides a sense of how well the model explains the variation in the data.

2. What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sumof Squares) in regression. Also mention the equation relating these three metrics with each other.

   In regression analysis, TSS, ESS and RSS are metrics that measure the goodness of fit of a model.

   TSS (Total Sum of Squares) is a measure of the total variance in the dependent variable. It represents the sum of the squared differences between the observed values and the mean of the dependent variable. The equation for TSS is:

   TSS = Σ(Yi - Ymean)^2

   ESS (Explained Sum of Squares) is a measure of the variance in the dependent variable that is explained by the independent variables in the model. It represents the sum of the squared differences between the predicted values and the mean of the dependent variable. The equation for ESS is:

   ESS = Σ(Ypred - Ymean)^2

   RSS (Residual Sum of Squares) is a measure of the variance in the dependent variable that is not explained by the independent variables in the model. It represents the sum of the squared differences between the observed values and the predicted values. The equation for RSS is:

   RSS = Σ(Yi - Ypred)^2

   Relationship between these three Metrics

   TSS = ESS + RSS

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

   Regularization is used to prevent overfitting, improve the generalization performance of a model, and make it more robust to noisy or irrelevant features in the data.

   The most common forms of regularization used in machine learning are L1 and L2 regularization, also known as Lasso and Ridge regularization, respectively.

4. What is Gini–impurity index?

The Gini impurity index is a measure of the disorder of a set of elements and is used to determine the best feature to split on in decision tree algorithms for classification.

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

Yes, unregularized decision trees are prone to overfitting.Unregularized decision trees allow the tree to grow very deep and complex, creating many branches and leaves. This can lead to the tree fitting the training data too closely, including capturing random noise in the data. When applied to new, unseen data, this overfitted tree may not perform well, leading to high prediction errors.Overfitting occurs because the unregularized decision tree algorithm places no restrictions on the size or shape of the tree, allowing it to fit the training data too closely. As a result, it becomes highly sensitive to the noise and outliers in the training data, which can lead to poor generalization performance on new data.Therefore, regularization techniques, such as pruning or setting a minimum number of samples required to split a node, are often used to prevent overfitting in decision trees.

6.What is an ensemble technique in machine learning?

Ensemble methods have been shown to be highly effective in various machine learning tasks, including classification, regression, and clustering. They are widely used in real-world applications, particularly in fields such as computer vision, natural language processing, and recommendation systems.

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

Bagging (Bootstrapped Aggregating) and Boosting are both ensemble techniques in machine learning that combine multiple individual models to form a more robust, composite model. However, they differ in how they generate the individual models and combine their predictions:

Bagging: Bagging involves training multiple instances of the same base model on different randomly selected subsets (bootstraps) of the training data. The goal is to reduce the variance of the model by training different models on different data and averaging their predictions. Bagging is best used with base models that have high variance, such as decision trees.

Boosting: Boosting involves training multiple instances of the base model sequentially, where each model focuses on correcting the mistakes made by the previous model. Boosting algorithms can adjust the weight of the samples in the training data, giving more importance to samples that are misclassified by previous models. The goal is to reduce the bias of the model by training models that focus on the difficult cases. Boosting is best used with base models that have high bias, such as linear regression.

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

The out-of-bag (OOB) error in random forests is a measure of the accuracy of the random forest model that can be used for model selection and hyperparameter tuning.

In random forests, the training data is divided into multiple subsets, and each tree in the forest is trained on a different subset of the data. This is known as bootstrapped aggregating, or "bagging." When training a random forest, some data points will not be included in the training set for any individual tree. These "out-of-bag" data points can be used to estimate the error of the random forest model without the need for cross-validation.

The OOB error is a convenient and fast way to estimate the performance of a random forest model without the need for a separate validation set or cross-validation. It can be used to compare different random forest models and select the best one, or to tune hyperparameters such as the number of trees and the maximum depth of each tree.

9. What is K-fold cross-validation?

K-fold cross-validation is a model evaluation technique used in machine learning to assess the performance of an algorithm. It involves dividing the dataset into K equal parts, or folds, and using K-1 folds for training and one fold for testing. This process is repeated K times, with each fold serving as the test set once. The average performance score across all K iterations is used as the performance metric for the model. This approach helps to ensure that the model is trained and tested on different data, reducing the chance of overfitting and improving the accuracy of the model evaluation.

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

Hyperparameter tuning is the process of adjusting the hyperparameters of a machine learning model to obtain the best performance on a given task. Hyperparameters are parameters that are set prior to training the model and cannot be learned from the data during training. They control the model's behavior and architecture, and include options such as the learning rate, number of hidden layers, and regularization strength

Hyperparameter tuning is done because the performance of a machine learning model is highly dependent on the choice of hyperparameters. The right hyperparameters can lead to better model performance and improved accuracy, while the wrong hyperparameters can result in underfitting or overfitting.

Hyperparameter tuning is typically performed by selecting a range of potential values for each hyperparameter, training a model using different combinations of hyperparameters, and evaluating the performance of each model on a validation set. The goal is to find the hyperparameters that result in the best performance, which is then used to train the final model.

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

A large learning rate in gradient descent can result in the following issues:Oscillation: If the learning rate is too large, the optimization process may oscillate back and forth across the minimum instead of settling into it, making it difficult to converge to a solution.Overshooting: The optimizer may overshoot the minimum and end up on the other side, leading to slow convergence or even divergence.Slow Convergence: Large learning rates can cause rapid progress at first, but eventually, the optimizer will converge too slowly or not at all.Unpredictable behavior: A large learning rate can cause the model parameters to jump around randomly, leading to unpredictable behavior and poor model performance.

Therefore, choosing an appropriate learning rate is important for ensuring the stability and convergence of the gradient descent optimization process. The learning rate can be chosen through a process called hyperparameter tuning, where multiple learning rates are tried and the one that gives the best performance is selected.

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

Logistic regression is a linear model and is not well suited for non-linear classification tasks. The logistic regression model a prediction based on a linear combination of the input features and their associated weights. If the relationship between the input features and the target variable is non-linear, this model may not be able to capture it accurately.

In such cases, more complex models like decision trees, random forests, or neural networks may be better suited to handle non-linear relationships. These models are capable of modeling non-linear relationships between the inputs and outputs by introducing non-linear transformations and interactions between the inputs.

However, in some cases, logistic regression can still be used for non-linear classification by transforming the input features. For example, by transforming the inputs into polynomial features, the logistic regression model can capture non-linear relationships between the inputs and outputs. However, this approach can result in a high-dimensional feature space, which may lead to overfitting.

13. Differentiate between Adaboost and Gradient Boosting.

AdaBoost and Gradient Boosting are both ensemble methods used for improving the performance of machine learning models. However, they differ in terms of how they generate the ensemble of models.

- **AdaBoost (Adaptive Boosting): AdaBoost is an iterative algorithm that adjusts the weights of the training examples at each iteration to give more importance to the examples that are misclassified by the current model. The final prediction**

is obtained by combining the weighted predictions of multiple weak models, with each model trying to correct the errors made by the previous models.

- **Gradient Boosting: Gradient Boosting is an iterative algorithm that trains a sequence of models to make a prediction. At each iteration, the model tries to minimize the residuals, which are the differences between the actual values and the predictions of the previous models. The final prediction is obtained by adding up the predictions of the individual models.**

In summary, AdaBoost emphasizes the importance of the misclassified examples in the training process, while Gradient Boosting tries to minimize the residuals of the previous models. Both methods have been shown to be effective for a wide range of tasks, but Gradient Boosting is generally considered to be a more powerful method as it can model more complex relationships between the inputs and outputs. However, it can also be more computationally expensive and prone to overfitting.

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

The bias-variance trade-off is a fundamental concept in machine learning that refers to the balance between two sources of error in a model: bias and variance.

- **Bias: Bias refers to the error introduced by assuming that the relationship between the inputs and outputs is too simple. A model with high bias is inflexible and may not fit the true relationship well, leading to underfitting and high training error.**
- **Variance: Variance refers to the error introduced by the model being too flexible and sensitive to the training data. A model with high variance is prone to overfitting and will perform well on the training data but poorly on new, unseen data.**

In machine learning, the goal is to find a model that has low bias and low variance, striking a balance between the two. A model with high bias may have low variance and vice versa. Finding the right balance is important for avoiding underfitting and overfitting, and for improving the generalization performance of the model on new, unseen data.

The bias-variance trade-off can be influenced by various factors, such as the complexity of the model, the amount and quality of the training data, and the regularization techniques used. Finding the right balance is an important aspect of model selection and optimization in machine learning.

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

Linear Kernel: The linear kernel is the simplest type of kernel and performs a simple dot product between two vectors to determine the similarity. It is used for linear classification problems.

RBF (Radial basis function) Kernel: The RBF kernel maps the input data into an infinite dimensional space and finds a hyperplane with maximum margin. This kernel is used in non-linear classification problems.

Polynomial Kernel: The polynomial kernel is a non-linear kernel that operates by raising the dot product of the two inputs to a given power (degree). It is used for non-linear classification problems.

The relationship between TSS, ESS, and RSS can be represented by the following equation:TSS = ESS + RSS

In other words, the total variance in the dependent variable is equal to the sum of the variance explained by the model and the variance not explained by the model. The smaller the RSS, the better the model fits the data.

6. 
7.
9.
10. What is the need of regularization in machine learning?
11. What is Gini–impurity index?
12. Are unregularized decision-trees prone to overfitting? If yes, why?
13. What is an ensemble technique in machine learning?
14. What is the difference between Bagging and Boosting techniques?
15. What is out-of-bag error in random forests?
16. What is K-fold cross-validation?
17. What is hyper parameter tuning in machine learning and why it is done?
18. What issues can occur if we have a large learning rate in Gradient Descent?
19. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
20. Differentiate between Adaboost and Gradient Boosting.
21. What is bias-variance trade off in machine learning?
22. Give short description each of Linear, RBF, Polynomial kernels used in SVM.