MACHINE LEARNING ASSIGNMENT - 5 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?

R-squared is better especially when comparing multiple complex models and it is easy to understand

1. 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.

TSS: It measure total variance in dependent variables

TSS=∑(yi​−yˉ​)2

where yi is the observed values and yˉ​ is the mean of the observed values.

ESS : It measures the portion of the total variance that is explained by the regression model. It shows how much of the variability in the dependent variable can be attributed to the independent variables.

ESS -∑ (y^- ˉy)2

Where y^ is predicted values

RSS: It show discrpency between predicted and observed values.

RSS-∑ (yi​−yˉ​)2

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

Regularization is needed for improving model generalization, controlling complexity, ensuring stability, aiding feature selection, and enhancing interpretability, ultimately leading to more for reliable machine learning models.

1. What is Gini–impurity index?

The Gini impurity index is a measure used in decision tree algorithms to assess the quality of a split at each node.

Gini(D)= 1− ∑C​pi2​

i= 1

here is D= number of data set

C = number of classes

Pi= is the proportion of instances in class i within the dataset D

.

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

Yes as it create more nodes which turns increase complexity in data and when we limit number of nodes the result may vary

1. What is an ensemble technique in machine learning?

Ensemble techniques in machine learning is which we combine multiple models to improve overall performance, stability. Tp achieve better accuracy than any single model could on its own.

1. What is the difference between Bagging and Boosting techniques?
2. Bagging aims to reduce variance and enhance model stability by training models independently on random subsets of data,

Boosting seeks to reduce bias and improve accuracy by training models sequentially and focusing on mistakes made by previous models

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

Out-of-bag error is a valuable feature of the Random Forest algorithm as it allows for an efficient and effective estimation of model performance. By utilizing the data not included in the training of each tree, OOB error provides a robust measure of how well the ensemble will generalize to new, unseen data.

1. What is K-fold cross-validation?

K-fold cross-validation is technique used to evaluate the performance of a machine learning model. It helps to ensure that the model is generalizing well to unseen data by providing a better estimate of its performance compared to a single train-test split.

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

Hyperparameter tuning in machine learning refers to the process of optimizing the hyperparameters of a model to improve its performance. Hyperparameters are the configuration settings that are not learned from the training data but are set before the training process begins. They control various aspects of the training process and the structure of the model itself.

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

Using a large learning rate in gradient descent can lead to several significant issues wihch negatively impact the training process and model performance. Here are the main problems associated with a large learning rate:

1 A large learning rate can overshoot the optimal point in the loss landscape.

2 If the learning rate is too high, the updates can result in the loss function increasing rather than decreasing.

3: High learning rates can introduce instability in the training process, where the loss fluctuates significantly from one iteration to the next and it may skip small but important data

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

logistic regression can be applied to datasets with non-linear boundaries, it has some limitations when dealing with non-linear data.

 Loistic regression assumes that the relationship between the features and the log-odds of the target variable is linear. This means it will create a linear decision boundary in the feature space.

For non-linear data, this assumption can lead to poor performance, as the model may fail to capture the underlying pattern

Underfitting: When applied to non-linear data, logistic regression may underfit the model, failing to adequately represent the relationship between features and the target. This results in high bias and low accuracy.

1. Differentiate between Adaboost and Gradient Boosting.

AdaBoost and Gradient Boosting are both popular ensemble techniques used to improve the performance of machine learning models, particularly decision trees. However, they have distinct approaches and characteristics. Here are the key differences:

**AdaBoost (Adaptive Boosting)**:

 **Focus on Misclassified Instances**: AdaBoost works by adjusting the weights of misclassified instances in each iteration. It gives more importance to instances that previous classifiers misclassified, allowing subsequent classifiers to focus on difficult cases.

**Sequential Learning**: It combines the outputs of multiple weak learners (often shallow decision trees) in a sequential manner, where each new model is added to correct the errors of the existing ensemble.

**Gradient Boosting**:

 **Gradient Descent Approach**: Gradient Boosting builds models sequentially, but instead of focusing solely on misclassified instances, it fits a new model to the difference between the actual and predicted values of the previous models.

**Optimization Objective**: It uses gradient descent to minimize a specified loss function, allowing it to optimize more complex objectives beyond simple classification.

**Models Type**

AdaBoost

Typically uses weak learners such as decision stumps (one-level decision trees). The simplicity of the base learners is a key characteristic of AdaBoost.

Gradient Boosting: Can use more complex models as base learners (e.g., deeper trees), giving it the ability to model intricate patterns in the data.

**Sensity to Noise and Outliers:**

AdaBoost: More sensitive to noise and outliers because it focuses on correcting misclassifications. Noisy instances can disproportionately affect the model since their

Gradient Boosting: Generally more robust to outliers, especially when appropriate loss functions (like Huber loss) are used. It minimizes overall prediction error rather than focusing solely on misclassifications.

**Performance and speed:**

AdaBoost :Generally faster to train than Gradient Boosting due to its simpler updating mechanism and typically fewer iterations.

Gradient Boosting :May take longer to train due to the need for fitting to residuals and potentially deeper trees, but it often yields better performance in terms of accuracy and flexibility.

In summary, AdaBoost and Gradient Boosting are both powerful ensemble techniques but operate on different principles. AdaBoost focuses on correcting misclassifications with adaptive weighting, while Gradient Boosting optimizes a loss function by fitting new models to the residuals. Each has its strengths and weaknesses, and the choice between them depends on the specific characteristics of the dataset and the problem at hand.

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

The bias-variance trade-off is a critical consideration in machine learning, influencing model selection, evaluation, and tuning. Striking the right balance between bias and variance is essential for building models that generalize well to new, unseen data, ultimately leading to better predictive performance.

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

Support Vector Machines (SVM) utilize kernels to transform data into higher-dimensional spaces, allowing for non-linear decision boundaries.