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?

Sol-R-squared is a better measure of goodness of fit in regression as it provides a normalised indicaton of how well the model's independent variables explain the variation in the dependent variable ,ranging from 0to1. R square & RSS are both measure to assess the goodness of fit of a regression model,but at different purposes.R square is directly related to hypothesis testing in the context of regression.RSS on the otherhand is a absolute measure that represents the sum of squared difference between observed and predicated values and it quantifies the overall error or residues in the model predictions.R square provide high level summary of the explanatory power of your model while RSS gives you detailed view of model's prediction accuracy.

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.

Sol- a)TSS represents the total variablitiy in the dependent variable(y).It is the sum of the squared difference between each observed Y values & the mean of all y values

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

b) ESS represent the variability in the dependent variable that is explained by the regression model, It is the sum of squared differences between predicated value from he regression model and the mean of the dependent variable.

$$ESS = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

C) RSS represent the unexplained variability in the dependent variable .
It is the sum of the squared difference between the observed dependent variable values and the predicted value from the regression model.

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

The relationship between

$$TSS = ESS + RSS$$

This equation highlights the total variability in the dependent variable (TSS) can be decomposed in to two components.the higher the proportion of the TSS explained by the model (higher ESS) the better in the models goodness fit.

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

Sol- Regularization in machine learning is employed to address several issues & improve the performance & generalization of models or in my terms I can say regularization in machine learning like adding a speed bumps to a road to prevent a car from going too fast and crashing.

Primary needs for regularization --

a) preventing overfitting - overfitting occurs when a model learns the training data too well,capturing noise or specific pattern that do not generalize to new ,unseen data .

b) handling multicollinearity

c) Feature selection

d) improving model stability

Regularization techniques, such as L1 (Laso) & L2 (Ridge) and combination of both play a crucial role .

4 What is Gini–impurity index?

Sol - The Gini impurtiy index is a measure of impurity used in decision tree algorithm for classification problems.It calculate the likelihood of misclassifying a randomly chosen element in a dataset based on the distribution of class labels.the formula for gini impurity index

$$Gini(p) = 1 - \sum_{i=1}^{c} p_i^2$$

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

Sol - yes unregularized decision tree are prone to overfitting , the primary reason for the susceptibility lies in the inherent nature of decision trees to create high complex structure that can be prefectly fit the training

data Decision trees aims to minimize impurity (eg Gini impurity or entropy) at each node ,leading to nodes that perfectly classify or nearly perfectly classify the training examples.

Key reasons:

a) memorization of noise

b) excessive complexity

C) lack of generalization constraints

6 What is an ensemble technique in machine learning?

Sol- An ensemble technique in machine learning involves combining predictions from a multiple individual models to create a more robust & accurate model.Methods like baggging(eg random forest) ,boosting & stacking aim to mitigate individuals model weakness, providing improved overall performance by leaveaging the diversity & collective wisdom of multiple models.

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

Sol- Bagging is the simplest way of combining predictions that belongs to the same type while boosting is a way of combining predictions that belong to different type.bagging aims to decrease variance, not bias whi;e boosting aims to decrease bias,not variance.bagging train multiple model indepently on different subset of the training data n model vote independently ,and the final predication is an average or majority vote . example random forest on the othe hand boosting train a sequence of models where each corrects error of the previous one.eg adaboost,gradient bosting,XG boost.

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

Sol- out of bag (oob) error also called out of bag estimate ,is a method of measuring the prediction error of a random forests, boosted decision trees , & other machine learning models utilizing bootstrap aggregating.In random forest algorithm ,the oob error is a way to estimate the models performance without the need for a separate validationn set.

9.What is K-fold cross-validation?

Sol- k fold cross validation is a technique for evaluating predictive models ,K fold cross validation is widely used technique in machine learning for assesing the performance & generalization ability of a model.

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

Sol- hyperparameter are configuration variables that are set before the training process of a model begins. They control the learning process itself, rather than being learned from the data.hyperparameter are often used to tune the performance of a model. It is done to find configuration of external settings for a machine learning models. It aims to optimize model performance,improve generalization to new data & avoid issues like overfitting or underfitting.

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

Sol - when the learning rate is too large , gradient descent can suffer from divergence . this means that weights increase exponentially, resulting in exploding gradients which can cause problems such as instabilities and overly high loss values.

To mitigate these issues, it is common practice to perform hyperparamter tuning and choose an appropriate learning rate.techniques such as learning rate scgedules, adaptive learning rates or using a smaller learning rate along with techniques like gradient clipping can help ensure stable and efficient convergence during the training of machine learning models.

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

Sol- Logistic regression is inherently designed for linear classification and may struggle with non - linear data . It assumes a linear decision boundary making it less effective for capturing cmplex relationship present in non-linear datasets. For non-linear data,more flexible models like support vector machines with non linear kernels or decision tree may be more suitable.

13 .Differentiate between Adaboost and Gradient Boosting.

Sol- adaboost

a) Assign weights to data points,focussing on misclassified ones.

b) Sequentially trains weak learners giving higher weights to misclassified loss function during training.

c) Can be sensitive to outliers & noisy data due to weight asignment.

d) Minimizes exponential loss function during training.

Gradient Boosting

a) Fits weak learners to minimize residuals(difference between predict & actual values)

b) sequential training without explicit data point weights.

c) minimizes the negative gradient of the loss function with respect to predicted values.

d) more robust to outliers & noisy data compared to adaboost.

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

Sol- The bias variance tradeoff in machine learning is the equilibrium between model simplicity(bias)

& flexibility (variance). High bias can lead to underfitting,ignoring patterns, while high variance can cause Overfitting capturing noise. Achieving the right balance is crucial for models to generalize well to new Data & perform effectively on both training & unseen datasets.

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

Sol -

Linear Kernel:

Description: The linear kernel computes the dot product between data points in the original feature space.
Use Case: Suitable for linearly separable data or when the number of features is large.

RBF (Radial Basis Function) Kernel:

Description: The RBF kernel measures the similarity between data points using a Gaussian function, allowing SVM to handle non-linear decision boundaries.
Use Case: Effective for capturing complex relationships and suitable for datasets with non-linear structures.

Polynomial Kernel:

Description: The polynomial kernel calculates the similarity between data points using polynomial functions, allowing SVM to capture polynomial relationships.
Use Case: Useful when the decision boundary is polynomial in nature, allowing SVM to handle curved decision boundaries.
In SVM, the choice of the kernel function depends on the nature of the data and the complexity of the decision boundary required for optimal classification.