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- R-squared is better measure of goodness of fit model in regression.because the proportion of variance in the dependent variable that is explained by the independent variables in the model. It ranges from 0 to 1, with a higher value indicating a better fit. R-squared is useful for comparing different models or for determining the proportion of the variability in the dependent variable that is explained by the model.

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.

Ans-TSS explained the how much variation there is in depenent variable.

TSS=∑(x-mean of x)^2

ESS --- The explained sum of squares (ESS) is the sum of the squares of the deviations of the predicted values from the mean value of a response variable, in a standard regression model .
It tells how much of the variation between observed data and predicted data is being explained by the model proposed. Mathematically, it is the sum of the squares of the difference between the predicted data and mean data.
Let $y_i = a + b_1x_{1i} + b_2x_{2i} + ... + \varepsilon_i$ is regression model, where

y_i is the i^{th} observation of the response variable

x_{ji} is the i^{th} observation of the j^{th} explanatory variable

a and b_i are coefficients

i indexes the observations from 1 to n

ε_i is the i^{th} value of the error term

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

Then

RSS------

The residual sum of squares (RSS) measures the level of variance in the error term, or residuals, of a regression model.

The smaller the residual sum of squares, the better your model fits your data; the greater the residual sum of squares, the poorer your model fits your data.

A value of zero means your model is a perfect fit.

$RSS = \sum^{n}_{i=1} (y^i - f(x_i))^2$

Where:
y_i = *the* i^{th} *value of the variable to be predicted*
$f(x_i)$ = *predicted value of* y_i
n = *upper limit of summation*

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

Ans-The need to regularize a model will tend to be less and less as you increase the number of samples that you want to train the model with or your reduce the model's complexity. However, the number of examples needed to train a model without (or with a very very small regularization effect) increases exponentially with the number of parameters and possibly some other factors inherit in a model

4. What is Gini–impurity index?

Ans--Gini Impurity tells us what is the probability of misclassifying an observation.

It measures the impurity of the nodes and is calculated as:

Gini Impurity=1-Gini

It only performs binary splits either yes or no, success or failure, and so on. So it will only split a node into two sub-nodes. These are the properties of Gini impurity.

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

Ans-Yes because they can keep growing deeper and more complex until they perfectly classify the training data. This can lead to the tree capturing noise in the data, rather than the underlying relationships, and thus performing poorly on new, unseen data. Regularization techniques such as pruning, setting a minimum number of samples required to split a node, or limiting the maximum depth of the tree can help mitigate overfitting in decision trees. Regularization techniques aim to simplify the tree and prevent it from becoming overly complex, thus improving its ability to generalize to new data.

6. What is an ensemble technique in machine learning?

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.

In learning models, noise, variance, and bias are the major sources of error. The ensemble methods in machine learning help minimize these error-causing factors, thereby ensuring the accuracy and stability of machine learning (ML) algorithms

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

Ans--Bagging

(i)Bagging is a learning approach that aids in enhancing the performance, execution, and precision of machine learning algorithms.

(ii)It is the easiest method of merging predictions that belong to the same type.

(iii)Here, every model has equal weight.

(iv)In bagging, each model is assembled independently.

(v)It helps in solving the over-fitting issue.

(vi)In the case of bagging, if the classifier is unstable, then we apply bagging.

Boosting

(i) Boosting is an approach that iteratively modifies the weight of observation based on the last classification.

(ii)It is a method of merging predictions that belong to different types.

(iii) Here, the weight of the models depends on their performance.

(iv)In boosting, the new models are impacted by the implementation of earlier built models.

(v) It helps in reducing the bias.

(vi)In the case of boosting, If the classifier is stable, then we apply boosting.

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

Ans-This approach utilizes the usage of bootstrapping in the random forest. Since the bootstrapping samples the data with the possibility of selecting one sample multiple times, it is very likely that we won't select all the samples from the original data set. Therefore, one smart decision would be

exploit somehow these unselected samples, called out-of-bag samples.

Correspondingly, the error achieved on these samples is called out-of-bag error. What we can do is to

use out-of-bag samples for each decision tree to measure its performance. This strategy provides reliable

results in comparison to other validation techniques such as train-test split or cross-validation.

9. What is K-fold cross-validation?

Ans--Cross validation is an evaluation method used in machine learning to find out how well your machine learning model can predict the outcome of unseen data. It is a method that is easy to comprehend, works well for a limited data sample and also offers an evaluation that is less biased, making it a popular choice.The data sample is split into 'k' number of smaller samples, hence the name: K-fold Cross Validation. You may also hear terms like four fold cross validation, or ten fold cross validation, which essentially means that the sample data is being split into four or ten smaller samples respectively.

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

Ans--Hyperparameter tuning is an essential part of controlling the behavior of a machine learning model. If we don't correctly tune our hyperparameters, our estimated model parameters produce suboptimal results, as they don't minimize the loss function. This means our model makes more errors.Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.

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

Ans-Learning rate to high like 0.5 or 0.7. In this case the major portion of gradient of current sample or current batch will be used for weight updation. Now for the next batch or sample again the new gradient will be generate which is different then the previous one and the major portion (as the learning rate is high as 0.5) is going to be used for weight updation. There are good chances that the weight updation caused by previous gradient may
adversely get affected by this current gradient (due to the chances that the gradient might be of opposite direction). And due to all this the learning wont happen and network keeps on learning from the recent examples rather than generalizing.

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

13. Differentiate between Adaboost and Gradient Boosting.

Ans-It identifies complex observations by huge residuals calculated in prior iterations
The trees with week learners are constructed using a greedy algorithm based on split points and purity scores. The trees are grown deeper with eight to thirty-two terminal nodes. The week learners should stay a week in terms of nodes, layers, leaf nodes, and splits
The classifiers are weighted precisely and their prediction capacity is constrained to learning rate and increasing accuracy
The shift is made by up-weighting the observations that are miscalculated prior
The trees are called decision stumps.
Every classifier has different weight assumptions to its final prediction that depend on the performance.
It gives values to classifiers by observing determined variance with data. Here all the week learners possess equal weight and it is usually fixed as the rate for learning which is too minimum in magnitude.
The exponential loss provides maximum weights for the samples which are fitted in worse conditions.

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

Ans-Bias in machine learning refers to the difference between a model's predictions and the actual distribution of the value it tries to predict. Models with high bias oversimplify the data distribution rule/function, resulting in high errors in both the training outcomes and test data analysis results.Bias is typically measured by evaluating the performance of a model on a training dataset. One common way to calculate bias is to use performance metrics such as mean squared error (MSE) or mean absolute error (MAE), which determine the difference between the predicted and real values of the training data.Bias is a systematic error that occurs due to incorrect assumptions in the machine learning process, leading to the misrepresentation of data distribution.The level of bias in a model is heavily influenced by the quality and quantity of training data involved. Using insufficient data will result in flawed predictions. At the same time, it can also result from the choice of an inappropriate model.

High-bias model features

Underfitting. High-bias models often underfit the data, meaning they oversimplify the solution based on generalization. As a result, the proposed distribution does not correspond to the actual distribution
Low training accuracy. The lack of proper processing of training data results in high training loss and low training accuracy
Oversimplification. The oversimplified nature of high-bias models limits their ability to identify complex features in the training data, making them inefficient for solving complicated problems.Variance stands in contrast to bias; it measures how much a distribution on several sets of data values differs from each other. The most common approach to measuring variance is by performing cross-validation experiments and looking at how the model performs on different random splits of your training data.

A model with a high level of variance depends heavily on the training data and, consequently,

has a limited ability to generalize to new, unseen figures. This can result in excellent

performance on training data but significantly higher error rates during model verification on

the test data.

High-variance model features

Low testing accuracy. Despite high accuracy on training data, high variance models tend to perform poorly on test data.

Overfitting. A high-variance model often leads to overfitting as it becomes too complex.
Overcomplexity. As researchers, we expect that increasing the complexity of a model will result in improved performance on both training and testing data sets. However, when a model becomes too complex and a simpler model may provide the same level of accuracy, it's better to choose the simpler one.

lexibility.

A complex model can learn complicated functions, which leads to higher variance. However, if the model

becomes too complex for the dataset, high variance can result in overfitting. Low variance indicates a

limited change in the target function in response to changes in the training data, while high variance means a significant difference.

How to reduce high variance?

The following methods can be used to overcome high variance:

Reducing the number of features in the model.
Replacing the current model with a simpler one.
Increasing the training data diversity to balance out the complexity of the model and the data structure.
Avoiding high-variance algorithms (support vector machines, decision trees, k-nearest neighbors, etc.) and opt for low-variance ones such as linear regression, logistic regression, and linear discriminant analysis.
Performing hyperparameter tuning to avoid overfitting.
Increasing regularization on inputs to decrease the complexity of the model and prevent overfitting.
Using a new model architecture. (Like with the high bias, this should be considered a last resort if other methods are not effective.)

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

Ans- Polynomial Kernel: It represents the similarity of vectors in the training set of data in a feature space over polynomials of the original variables used in the kernel.

Linear Kernel:it is used when data is linearly separable.

Gaussian Kernel Radial Basis Function (RBF): it is adding radial basis method to improve the transformation.