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?**

Residual Sum of Squares (RSS) is a statistical technique used in regression analysis to determine the dispersion of data points. In a regression analysis, the goal is to determine how well a data series can be fitted to a function that might help to explain how the data series was generated. The sum of squares is used as a mathematical way to find the function that best fits (varies least) from the data.

The RSS measures the amount of error remaining between the regression function and the data set after the model has been run. A smaller RSS figure represents a regression function that is well-fit to the data.

R-squared (R2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. Whereas correlation explains the strength of the relationship between an independent and dependent variable, R-squared explains to what extent the variance of one variable explains the variance of the second variable. So, if the R2 of a model is 0.50, then approximately half of the observed variation can be explained by the model's inputs.

RSS VS R2 which one to choose-- Residual Sum of Squares (RSS), when calculated, doesn't explain anything. It is merely a number. You must use this number RSS to divide by SST to make some sense out of it.

Where the R squared is the (RSS/SST), i.e the absolute amount of variation as a proportion of total variation.

R-squared (R2) is better.

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.**

TSS (Total Sum of Squares)- The sum of squares total or Total Sum of Squares , denoted SST or TSS, is the squared differences between the observed dependent variable and its mean.

ESS (Explained Sum of Squares)- sum of squares due to regression, or SSR. It is the sum of the differences between the predicted value and the mean of the dependent variable. Think of it as a measure that describes how well our line fits the data.
RSS (Residual Sum of Squares)- The residual sum of squares tells how much of the dependent variable's variation your model **did not explain**. It is the sum of the squared differences between the actual Y and the predicted Y.

TSS = ESS + RSS

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

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.
Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

Technique in regularization-Lasso and Ridge

4. **What is Gini–impurity index?**

Gini Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree. More precisely, the Gini Impurity of a dataset is a number between 0-0.5, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset.
Gini impurity index measures the diversity in a set. Let's say, for example, that we have a bag full of balls of several colors. A bag where all the balls have the same color, has a very low Gini impurity index (in fact, it is zero). A bag where all the balls have different colors has a very high Gini impurity index.

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

Yes,

- Overfitting Due to Presence of Noise Mislabeled instances may contradict the class labels of other similar records.
- Overfitting Due to Lack of Representative Instances Lack of representative instances in the training data can prevent refinement of the learning algorithm.

6. **What is an ensemble technique in machine learning?**

Ensemble is a machine learning concept, in which several models are trained using machine learning algorithms. It combines low performing classifiers (also called as weak learners or base learner) and combine individual model prediction for the final prediction.

Ensemble techniques are classified into three types:

- Bagging- In bagging every model is given equal preference instead of depending on one model, it runs the data through multiple models in parallel, and average them out as model's final output.
  Bagging reduces the variance.
- Boosting- is completely opposite of bagging, if one model predicts data more correctly than the other, then higher weightage should be given to this model over the other. Also, the model should attempt to reduce bias. These concepts are applied in the second ensemble method that we are going to learn, that is Boosting.
  Boosting reduces bias.
- **Stacking**

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

Bagging	Boosting
1. Aim to decrease variance.	1.Aim to decrease bias, not variance.
2. Each model receives equal weight.	2.Models are weighted according to their performance.
3. In this base classifiers are trained parallelly.	3.In this base classifiers are trained sequentially.
4. Example: The Random forest model uses Bagging techniques.	4.Example: The AdaBoost uses Boosting techniques.

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

K-fold cross-validation is defined as a method for estimating the performance of a model on unseen data. This technique is recommended to be used when the data is scarce and there is an ask to get a good estimate of training and generalization error thereby understanding the aspects such as underfitting and overfitting. This technique is used for hyperparameter tuning such that the model with the most optimal value of hyperparameters can be trained. It is a resampling technique without replacement. The advantage of this approach is that each example is used for training and validation (as part of a test fold) exactly once. This yields a lower-variance estimate of the model performance than the holdout method. As mentioned earlier, this technique is used because it helps to avoid overfitting, which can occur when a model is trained using all of the data. By using k-fold cross-validation, we are able to "test" the model on k different data sets, which helps to ensure that the model is generalizable.

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

Parameters which define the model architecture are referred to as **hyperparameters** and thus this process of searching for the ideal model architecture is referred to as *hyperparameter tuning*.

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

Gradient Descent is a simple optimization technique that could be used in many machine learning problems. It involves reducing the cost function. The cost function is the relation between calculated output and actual output.
If learning rate is too large, gradient descent can overshoot the minimum. It may fail to converge and even diverge. If learning rate is too small, gradient descent can be slow

12. **Can we use Logistic Regression for classification of Non-Linear Data? If not, why?**
No
Non-linear problems can't be solved with logistic regression because it has a linear decision surface.

13. **Differentiate between Adaboost and Gradient Boosting?**

AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem. This makes Gradient Boosting more flexible than AdaBoost.

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

If the algorithm is too simple (hypothesis with linear eq.) then it may be on high bias and low variance condition and thus is error-prone. If algorithms fit too complex (hypothesis with high degree eq.) then it may be on high variance and low bias. In the latter condition, the new entries will not perform well. Well, there is something between both of these conditions, known as Trade-off or Bias Variance Trade-off.

15. **Give short description each of Linear, RBF, Polynomial kernels used in SVM.**
Kernels or kernel methods (also called Kernel functions) are sets of different types of algorithms that are being used for pattern analysis. They are used to solve a non-linear problem by using a linear classifier. Kernels Methods are employed in SVM (Support Vector Machines) which are used in classification and regression problems.

Linear SVM: Linear SVM is used for **linearly** separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

RBF: It is used to perform transformation when there is no prior knowledge about data. most widely used kernels due to its similarity to the Gaussian distribution. The RBF kernel function for two points X_1 and X_2 computes the similarity or how close they are to each other.

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.