MACHINE LEARNING – 5

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.. RSS is the sum of the squares of the errors made by the model on each data point. So, RSS depend upon the number of data-points in the data. If dataset is large, then naturally it will have large RSS because RSS is the sum of squares of the errors made by all the data points. While on the other hand R-squared is given as follows:

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}$$

Where,

y i = True value of the dependent variable

y ˆ = predicted value of the dependent variable

y = mean value of dependent variable

So, R-squared does not depend upon the number of data-points in the data, rather it depends only on the quality of the fit of the curve on the data, while RSS depends on both the quality of fit and also the number of data points in the data.

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

Ans. While training the machine learning model the model can easily be overfitted and underfitted. To remove this, we use regularization in machine learning to properly fit the model onto our test set.

We can say that regularization technique helps to reduce the chances of overfitting and underfitting and helps to get the optimal model.

3. What is Gini–impurity index?

Ans.

Gini impurity is a function that determines how well a decision tree was split. Basically, it helps us to determine which splitter is best so that we can build a pure decision tree. Gini impurity ranges values from 0 to 0.5. It is one of the methods of selecting the best splitter.

4. What is an ensemble technique in machine learning?

Ans. Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Ensemble methods usually produces more accurate solutions than a single model would. Voting and averaging are two of the easiest ensemble methods. They are both easy to understand and implement. Voting is used for classification and averaging is used for regression.

5. What is the difference between Bagging and Boosting technique?

	Bagging	**Boosting**
1.	The simplest way of combining predictions that belong to the same type.	A way of combining predictions that belong to the different types.
2.	Aim to decrease variance, not bias.	Aim to decrease bias, not variance.
3.	Each model receives equal weight.	Models are weighted according to their performance.
4.	Each model is built independently.	New models are influenced by the performance of previously built models.
5.	Different training data subsets are selected using row sampling with replacement and random sampling methods from the entire training dataset.	Every new subset contains the elements that were misclassified by previous models.
6.	Bagging tries to solve the over-fitting problem.	Boosting tries to reduce bias.
7.	If the classifier is unstable (high variance), then apply bagging.	If the classifier is stable and simple (high bias) the apply

6. **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 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.

7. **What is K-fold cross-validation?**

Ans. K-fold Cross-Validation is when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data. K refers to the number of groups the data sample is split into. For example, if you see that the k-value is 5, we can call this a 5-fold cross-validation. Each fold is used as a testing set at one point in the process.

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

Ans. Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while 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.

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

Ans. Learning Rate is the hyperparameter that determines the steps the gradient descent algorithm takes. Gradient Descent is too sensitive to the learning rate. If it is too big, the algorithm may bypass the local minimum and overshoot.

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

Ans. Non-linear problems can't be solved with logistic regression because it has a linear decision surface. The decision boundary is a line or a plane that separates the target variables into different classes that can be either linear or nonlinear. In the case of a Logistic Regression model, the decision boundary is a straight line.

11. Differentiate between Adaboost and Gradient Boosting.

Ans. Adaboost increases the performance of all the available machine learning algorithms and it is used to deal with weak learners. It gains accuracy just above the arbitrary chances of classifying the problem. The adaptable and most used algorithm in AdaBoost is decision trees with a single level. The gradient boosting depends on the intuition which is the next suitable possible model, when get combined with prior models that minimize the cumulative predicted errors. The crucial idea of gradient boosting is to fix the targeted outcomes for the next model to reduce the error.

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

Ans. While building the machine learning model, it is really important to take care of bias and variance in order to avoid overfitting and underfitting in the model. If the model is very simple with fewer parameters, it may have low variance and high bias. Whereas, if the model has a large number of parameters, it will have high variance and low bias. So, it is required to make a balance between bias and variance errors, and this balance between the bias error and variance error is known as the Bias-Variance trade-off.

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

Ans. Linear kernels - Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are a Large number of Features in a particular Data Set.

Polynomial kernels- 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.

RBF kernel - When the data set is linearly inseparable or in other words, the data set is non-linear, it is recommended to use kernel functions such as RBF.

14. 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.

Ans.* TSS is total sum of squares. It is equal to the variance of the data.

- ESS is called the explained sum of squares. It is the variance of the data which has been explained by the model. The model is able to explain this much variance of the data. To refer the formula please look above in the image.

- RSS is called the residual sum of squares. It is equal to the sum of squares of all the errors or residuals made by the model on the data.

The relationship between tss, rss and ess is given as: TSS = ESS + RSS

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

Ans. Unregularized decision-trees are highly prone to overfitting. If we do not restrict the depth up to which a tree can be grown or control it in any other way, the decision tree will most likely learn each and every data point in the training dataset. So, it will learn the training data patterns too closely and when it will be tested on unseen data, it will most likely perform poorly. So, in order to solve this problem, we regularize decision trees by a number of ways, either by controlling the depth of the tree, or controlling the maximum number of leaves tree can have etc.