MACHINE LEARNING

Ans 1 :- The Residual Sum of Squares (RSS) is a statistical technique used to measure the amount of Variance in a data set. 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. Where as R-Squared is statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other word, R-squared shows how well the data fit the regression model the goodness of fit. A higher R-squared value indicates a higher amount of variability being explained by our model and vice-versa. If we had a really low RSS value, it would mean that the regression line was very close to the actual points. This means the independent variables explain the majority of variation in the target variable. The residual sum of squares is the absolute amount of explained variation, whereas R- squared is the absolute amount of variation as a proportion of total variation.

Ans 2 :- • TSS :- TSS is the sum of square of difference of each data point from the mean value of all the values of Target Variable (y). Here, the line with intercept ('c' in y = mx + c) equal to y mean. It means that this line does not include any influence of independent variable.

• ESS :- Earth system science (ESS). Is a survey that focus on the critical components in Earth science, including atmospheric , Ocean , Seismology and biosphere and cover AI/ML applications to statistical downscaling and forecasting problems.

• RSS :- Residual sum of squares (RSS) measures the level of variance in the error team 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.

Ans 3:- Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it. Sometimes the machine learning model performs well with the training data but does not perform well with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique .This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model. It mainly regularizes or reduces the coefficient of features toward zero. In simple words. 'In regularization technique, we reduce the magnitude of the features by keeping the same number of features'.

Ans 4:- 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. For example :- say you want to build a classifier that determines if someone will default on their credit card. You have some labelled data with features, such as bins for age, income , credit rating and wheather or not each person is a student. To find the best feature for the first split of the tree – the root node – you could calculate how poorly each feature divided the data into the correct class, default ("Yes") or not a default ("No"). this calculation would measure the "Impurity" of the split and the feature with the lowest impurity would determine the best feature for splitting the current node . this process would continue for each subsequent node using the remaining features.

Ans 5 :- One of the limitations of decision tree is that they are largely unstable compared to other decision predictors. A small change in the data can result in a major change in the structure of the decision tree. Decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample events that meet the previous assumptions. Overfitting affects the accuracy of predictions made from the samples which are not part of the training set. One can build a perfect decision tree model on the training data with 100% accuracy, but with significantly low accuracy on test data.

Ans 6 :- Ensemble methods are techniques that create multiple models combine them to produce improved results. Ensemble methods usually produces more Accurate solutions than a single model would. This has been the

case in a number of machine learning competitions, where the winning solutions used ensemble methods. The main challenge is not to obtain base models which make different kinds of error. For example:- if ensembles are used for classification , high accuracy can be accomplished if different base models misclassify different training examples, even if the base classifier accuracy is low.

Types of Ensemble Classifier :--

• Bagging :- Bagging (Bootstrap Aggregation) is used to reduce the variance within a noisy dataset. In bagging a random sample of data in a training set is selected with replacement meaning that the individual data points can be chosen more than once.

• Random Forest :- It is an extention over bagging. Each classifier in the ensemble is a decision tree classifier and is generated using a random selection of attributes at each node to determine the split. during classification, each tree votes and the most popular class is returned.

Ans 7 :- Difference between Bagging and Boasting :-

• Bagging :- Bagging attempts to tackle the over-fitting issue. In bagging every model receives an equal weight. Every model is constructed independently. Various training data subsets are randomly drawn with replacement from the whole training dataset. if the classifier is unstable (high variance),then we need to apply bagging.it is the easiest way of connecting predictions that belong to the same type.

• Boosting :- Boosting tries to reduce bias.in boosting models are weighted by their performance. New models are affected by the performance of the previously developed model. each new subset contains the components that were misclassified by previous models. If the classifier is steady and straightforward (high bias), then we to apply boosting.it is away of connecting predictions that belong to the different types.

Ans 8 :- *Out—of—bag (OOB) error , it is also called as Out—of—bad estimate is a method of measuring the prediction error of random forests, boosted decision trees and other machine learning models utilizing Bootstrap aggregating (bagging). Bagging uses subsampling with replacement to create training samples for the model to learn. OOB error is the mean prediction error on each training sample x. using only the trees that did not have x in their bootstrap sample. A prediction made for an observation in the original data set using only base learners not trained on this particular observation. These predictions are not prone to overfitting as each prediction is only made by learners that did not use the observation for training.*

Ans 9 :- *Cross- validation is a statistical method used to estimate the skill of machine learning models. It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is Bagging Boosting The simplest way of combining predictions that belong to the same type. Each model receives equal weight. A way of combining predictions that belong to the different types. Models are weighted according to their performance Each model is built independently New models are influenced by the performance of previously built models. Bagging tries to solve the over-fitting problem. Boosting tries to reduce bias. If the classifier is instable (high variance) then apply bagging If the classifier is stable and simple (high bias) then apply boosting. Aim to decrease variance, not Bias. Aim to decrease bias, not variance. In Bagging base classifiers are trained parallelly. In Boosting base classifier are trained sequentially. For example: the random forest model uses Bagging For example: the AdaBoost uses Boosting easy to understand, implement and results in skill estimates that generally have alower bias than other methods. Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called K refers to the number of groups that a given data sample is to be split into. As such, procedure is often called as K-fold-Cross Validation. When a specific value for K is chosen, it may be used in place of K in the reference to the model, such as K=5 becoming 5-fold-cross-validation. 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 a 5-fold-cross-validation.

Ans 10 :- In Machine learning, we need to differentiate between parameters and hyperparameters. A learning algorithm learns or estimates model parameters for the given data set, then continues updating these values as it continues to learn. After learning is complete, these parameters become part of the model. Hyperparameters are specific to the algorithms itself, so we can't calculate their values from the data. We use hyperparameters to calculate the model parameter.Different hyperparameter values produce different model parameter values for a given dataset. Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithms to any data set. That combination of hyperparameter maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors. Note that the learning algorithm optimizes the loss based on the input data and tries to find an optimal solution within the given setting. However, hyperparameters describe this setting exactly. 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. In practice, key indicators like the accuracy or the confusion matrix will be worse.

Ans 11 :- Learning rate is one such hyper-parameter that defines the adjustment in the weights of our network with respect to the loss gradient descent. It determines how fast or slow we will move towards the optimal weights. The Gradient Descent Algorithms estimates the weights of the model in many iterations by minimizing a cost function at every step. In order for Gradient Descent to work, we must set the learning rate to an appropriate value. This parameter determines how fast or slow we will move towards the optimal weights. If the learning rate is very large we will skip the optimal solution.If it is too many iterations to converge to the best values. So using a good learning rate is crucial. When the learning rate is to large descent can inadvertently increase rather than decrease the training error. When the learning rate is too small, training is not only slower, but may become

permanently stuck with a high training error. However, if the learning rate is set too large it can cause undesirable divergent behaviour in your loss function.

Ans 12 :- Logistic Regression is neither linear nor is it a classifier. The idea of a decision boundary has little to do with logistic regression, which is instead a direct probability estimation method that separates predictions from decision. Logistic regression has traditionally been used as a linear classifier. When the classes can be separated in the feature space by linear boundaries may be non-linear. No, we cannot use Logistic Regression for classification of Non-Linear Data as logistic regression only forms linear decision surface. And logistic regression is indeed non linear in terms of odds and probability, however it is linear in terms of Log Odds.

Ans 13 :- Diffrence between AdaBoost and GradientBoost:-

• AdaBoost :- In AdaBoost "shortcomings" are identified by high-weight data points. Exponential loss of AdaBoost gives more weights for those samples fitted worse. It is considered as a special case of Gradientboost in terms of loss function, in which exponential losses. shift is done by up-weighting observations that were misclassified before.

• GradientBoost :- Gradient boost "shortcomings"are identified by gradients. It identifies difficult observations by large residuals computed in the previous iterations. It further dissect error components to bring in more explanation.it concepts of gradients are more general in nature.

Ans 14 :- In statistics and machine learning, the bias-variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters. This is the conflict in trying to simultaneously minimize these two sources of error that prevent supervised machine learning algorithms from generalizing beyond the training set. If our model is too simple and has very few parameters then it may have high bias and low variance. But if our model has large number of parameters then it's going to have high variance and low bias. When we modify the machine learning algorithm to better fit a given dataset, it will in

turn lead to low bias and but will increase the variance. This way, the model will fit with the dataset while increasing the chances of inaccurate predictions.

Ans 15 :- • Linear:- The linear model is one of the simplest models in machine learning, but linear models are the building blocks for deep neural networks. There are two main classes in supervised machine learning problems, regression and classification. In reversal, the target value is the actual value.

•RBF: Radial Basis Functions (RBF) are real-valued functions that use supervised machine learning to perform as a non-linear classifier. Its value depends on the distance between the input and a certain fixed point.

• Polynomial kernels used in SVM: In machine learning polynomial kernel is a kernel function commonly used with Support Vector Machines (SVM) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models.

.