

MACHINE LEARNING

- 1) A residual sum of squares (RSS), also known as the sum of squared residuals (SSR), is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model. R² (R Squared) is another statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable in a regression model. R-Squared is the better measure of goodness of fit compared to RSS. 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.
- 2) The sum of squares total, denoted SST, is the squared differences between the observed dependent variable and its mean; for i in n $(y(i) - y(\text{mean}))^2$
The second term is the 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 for i in n $(\hat{y}(i) - y(\text{mean}))^2$
The last term is the sum of squares error, or SSE. The error is the difference between the observed value and the predicted value. We want usually to minimize the error. for i in n $(e(i))^2$

The relation between the above 3 could be linearly expressed as:

$$\text{TSS} = \text{RSS} + \text{ESS}$$

- 3) Regularization constraints or shrinks the coefficient towards zero, it is used to prevent Overfitting in ML Algos.
There are two types of regularization Lasso (L1 Norm) & Ridge (L2 Norm)
- 4) Gini Index, also known as Gini impurity, calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly. If all the elements are linked with a single class then it can be called pure.

Let's perceive the criterion of the Gini Index, like the properties of entropy, the Gini index varies between values 0 and 1, where 0 expresses the purity of classification. This measure is calculated where the modeling contains Tree Algorithms like Decision Tress or random forest.

- 5) Yes, decision tree is prone to overfitting. But unlike other algorithms decision tree does not use regularization to fight against overfitting. Instead it uses pruning. There are mainly two types of pruning performed:

Pre-pruning that stops growing the tree earlier, before it perfectly classifies training Set.

Post-pruning that allows the tree to perfectly classify the training set, and then post prune the tree.

6)

Ensemble techniques are the algorithms created combining multiple weak learners to a strong learning model. Random Forest, XG Boosts, Gradient Boosting are some examples of ensemble learning techniques. These are 2 types of Ensemble techniques, Bagging and Boosting.

7)

Bagging, which is also known as bootstrap aggregating sits on top of the majority voting principle, every model receives an equal weight; objective to decrease variance not bias.

Boosting is another ensemble procedure to make a collection of predictors. In other words, we fit consecutive trees, models are weighted by their performance, the objective is decrease bias not variance.

8)

Out of sample is a technique to verify the performance of a bootstrapping model without having to use a validation set. This is an advantage if:

Your data set is too small to split into training, validation and test.

Gives a second validation on the model allowing.

9)

K Fold cross validation means training and testing with different subset of the training and testing data so that the model won't be biased over some cords in the dataset (avoid overfitting). The K in K fold is the integer defining how many times does the subset should be created and trained and tested. For example a 5 Fold cross validation will create 5 subsets in both training and testing dataset, train

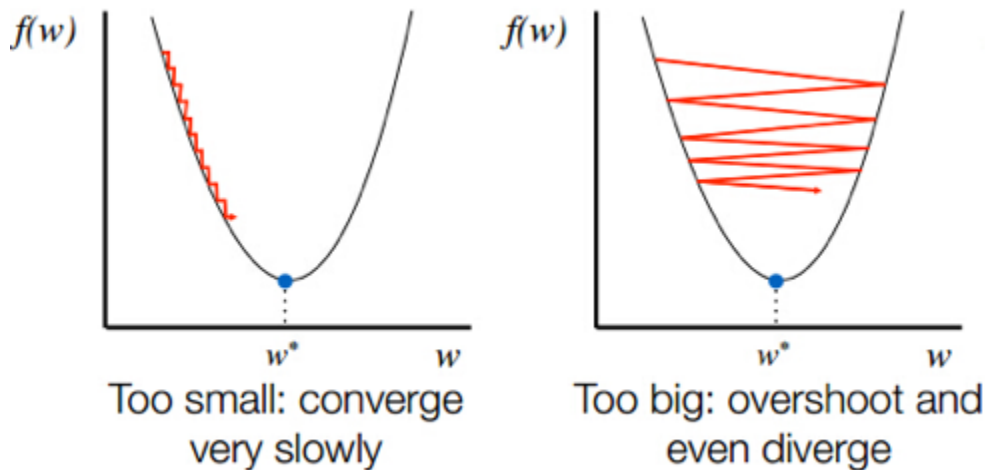
and predict are output 5 accuracy values. Averaging those values would gives us a grater idea of how good the model is.

10)

Hyper parameters are the parameters of the model algorithms which are to be tunes in order to get maximum accuracy from the machine learning model.

11)

When the learning rate is too large, then gradient descent will overshoot the minimum point and will ultimately fail to converge.



12)

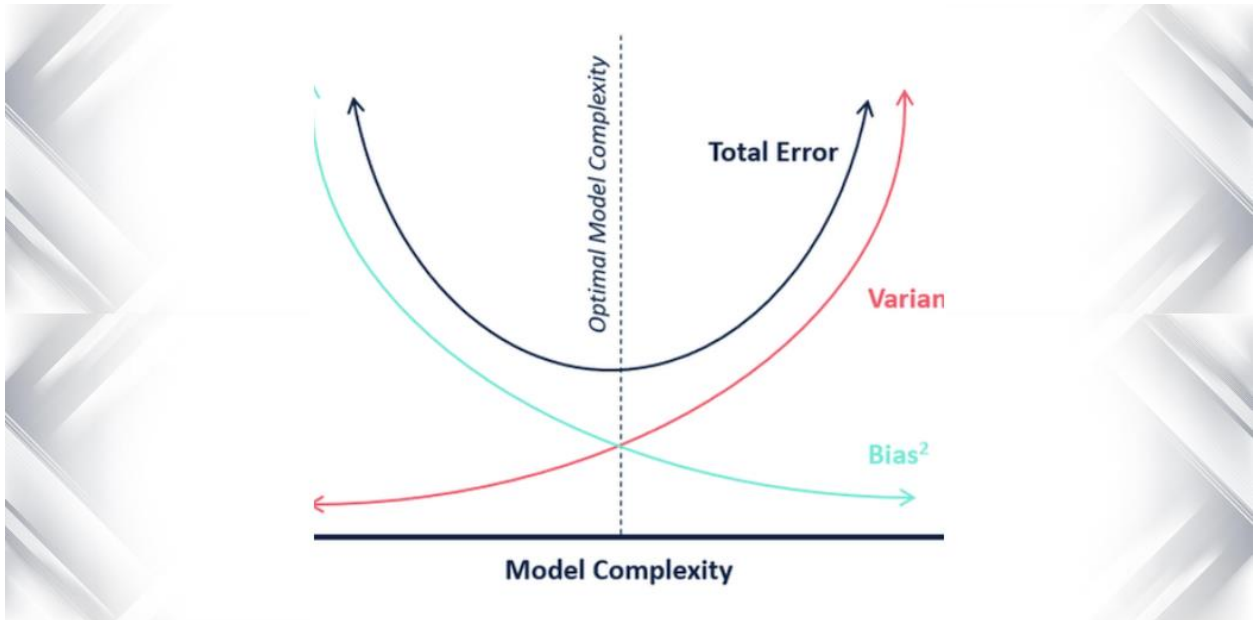
Logistic Regression has traditionally been used as a linear classifier, i.e. when the classes can be separated in the feature space by linear boundaries.

13)

Gradient boosting defies boosting as a numerical optimization problem where the objective is to minimize the loss function of the model by adding weak learners using gradient descent. Whereas, method focuses on training upon misclassified observations. Alters the distribution of the training dataset to increase weights on sample observations that are difficult to classify.

14)

There is a tradeoff between a model's ability to minimize bias and variance. Understanding these two types of error can help us diagnose model results and avoid the mistake of over- or under-fitting. This is known as bias-variance tradeoff.



15)

SVM also known as Support Vector Machine is a supervised machine learning algorithm which can be used for both classification or regression challenges. SVM uses different kernels for different types of questions.

A linear kernel allows you to use linear functions, which are really impoverished. As you increase the order of the polynomial kernel, the size of the function class increases. In the polynomial kernel, we simply calculate the dot product by increasing the power of the kernel. Gaussian RBF (Radial Basis Function) is another popular Kernel method used in SVM models for more. RBF kernel is a function whose value depends on the distance from the origin or from some point.

