**NAME OF THE PROJECT**

**Machine learning  
Submitted by:**

**SAUNAK MUKHERJEE**



1. R-squared is a better measure of goodness of fit model in regression.

a higher r-squared indicates more variability is explained by the model, that’s why it is better.

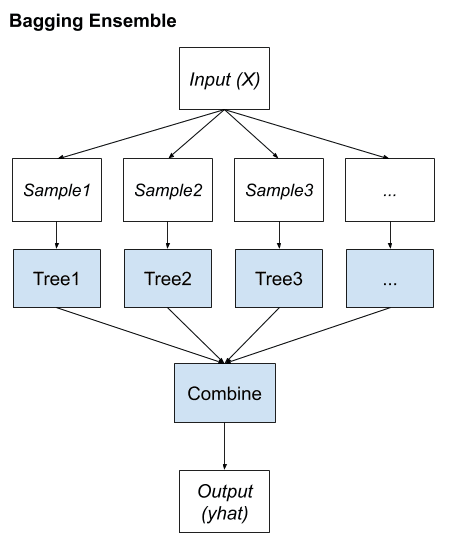
1. In the statistical data analysis the total sum of squares (TSS or SST) is a quantity that appears as part of a standard way of presenting results of such analyses

 ESS is Explained Sum of Squares

TSS = ESS + RSS

1. Regularization techniques which used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.
2. Gini impurity which is a function that determines how well a decision tree was split. It also 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

6 Ensemble learning helps improve machine learning results by combining several models. This approach allows the production of better predictive performance compared to a single model.



1. 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
2. Hyperparameter tuning consists of finding a set of optimal hyperparameter values which is 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.

14. The bias variance tradeoff is the tradeoff which happens between bias (error made by the model) and

variance (how much the model changes with change in training data) when the model complexity

changes. If the model is too simple the model makes too much errors and its predictions becomes

inaccurate, so when we decrease the model complexity the bias (or errors) increases but the variance

decreases (that is the model does not change much with change in training data). On the other hand,

if model is too complex the bias although becomes low but the variance increases. So, there is tradeoff

between bias and variance. So, we always have to find that sweet spot where the model does not

have much bias neither much high variance.

15. The SVM uses kernel functions to transform the data from one set of dimensions to another set of

dimensions so that the decision boundary in the resultant space is simpler than the decision boundary

in original space. Now, the kernel to be used depend upon the nature of the data we have.

• Linear kernel will be used if the original data is linearly separable.

• Polynomial kernel is used when the data is the form of polynomial of some degree n.

• RBF is used when the data follows some complex pattern which is neither linear nor

polynomial.

11. The two main issues which can occur if we perform gradient descent with large learning rate are:

• The gradient descent algorithm can diverge from the optimal solution if we try out a very large

learning rate. The algorithm can go away from the optimal solution if we have a very large learning

rate.

• The gradient may simply keep oscillating around the optimal solution if the learning rate is high,

and it will not settle at the optimal solution.

12. We cannot use Logistic Regression for classification of Non-Linear Data because the decision

boundary produced by logistic regression is linear and if we have nonlinear data where we have

nonlinear decision boundaries then if we try to use the logistic regression it will perform poor on the

data, as the decision boundary in the data is nonlinear.