

## MACHINE LEARNING-7

### Answer of Following Questions-

Ans 1-A) GridSearchCV() B) RandomizedCV()

Ans 2- A) Random forest

Ans 3-A) The regularization will increase

Ans 4- D) None of the above

Ans 5-D)None of the above

Ans 6-C) Both of them

Ans 7-B) Bias will decrease, Variance increase

Ans 8-B) model is overfitting

Ans 9-Gini is 0.48 and Entropy is 0.96

Ans 10-

- 1. Complexity: Random Forest creates a lot of trees (unlike only one tree in case of decision tree) and combines their outputs. ...
- 2. Longer Training Period: Random Forest require much more time to train as compared to decision trees as it generates a lot of trees (instead of one tree in case of decision tree) and makes decision on the majority of votes.

Ans 11- So, in cases where all the columns have a significant difference in their scales, are needed to be modified in such a way that all those values fall into the same scale. This process is called Scaling. There are two most common techniques of how to scale columns of Pandas dataframe – **Min-Max Normalization and Standardization**

.

Ans 12- Every time we train a deep learning model, or any neural network for that matter, we're using **gradient descent** (with backpropagation). We use it to *minimize a loss* by *updating the parameters/weights* of the model.

The parameter update depends on two values: a gradient and a **learning rate**. The learning rate gives you control of how big (or small) the updates are going to be. A **bigger learning rate** means **bigger updates** and, hopefully, a model that **learns faster**.

**Ans 13-** An imbalanced classification problem is an example of a classification problem where the distribution of examples across the known classes is biased or skewed. The distribution can vary from a slight bias to a severe imbalance where there is one example in the minority class for hundreds, thousands, or millions of examples in the majority class or classes.

Imbalanced classifications pose a challenge for predictive modeling as most of the machine learning algorithms used for classification were designed around the assumption of an equal number of examples for each class. This results in models that have poor predictive performance, specifically for the minority class. This is a problem because typically, the minority class is more important and therefore the problem is more sensitive to classification errors for the minority class than the majority class

**Ans 14-** The F-score (also known as the F1 score or F-measure) is a metric used to evaluate the performance of a Machine Learning model. It combines precision and recall into a single score. F-measure formula: **F-score =  $2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$**

**Ans 15-** In Data science and machine learning the methods like `fit()`, `transform()`, and `fit_transform()` provided by the [scikit-learn](https://scikit-learn.org/) package are one of the vital tools that are extensively used in data preprocessing and model fitting. The task here is to discuss what is the difference between **`fit()`**, **`transform()`**, and **`fit_transform()`** and how they are implemented using in-built functions that come with this package.

- The **`fit(data)`** method is used to compute the mean and std dev for a given feature to be used further for scaling.
- The **`transform(data)`** method is used to perform scaling using mean and std dev calculated using the `.fit()` method.
- The **`fit_transform()`** method does both fits and transform