

## **MACHINE LEARNING – WORKSHEET 7**

**In Q1 to Q8, only one option is correct, Choose the correct option:**

**1. Which of the following in sklearn library is used for hyper parameter tuning?**

**Ans:- A) GridSearchCV()**

**2. In which of the below ensemble techniques trees are trained in parallel?**

**Ans:- A) Random forest**

**3. In machine learning, if in the below line of code:**

**`sklearn.svm.SVC (C=1.0, kernel='rbf', degree=3)`**

**we increasing the C hyper parameter, what will happen?**

**linear**

**Ans:- B) The regularization will decrease**

**4. Check the below line of code and answer the following questions:**

**`sklearn.tree.DecisionTreeClassifier(*, criterion='gini', splitter='best', max_depth=None, min_samples_split=2)`**

**Which of the following is true regarding max\_depth hyper parameter?**

**Ans:- C) both A & B**

**5. Which of the following is true regarding Random Forests?**

**Ans:- A) It's an ensemble of weak learners.**

**6. What can be the disadvantage if the learning rate is very high in gradient descent?**

**them**

**Ans:- C) Both of them**

**7. As the model complexity increases, what will happen?**

**Ans:- B) Bias will decrease, Variance increase**

**8. Suppose I have a linear regression model which is performing as follows:**

**Ans:- B) model is overfitting**

**Q9 to Q15 are subjective answer type questions, Answer them briefly.**

**9. Suppose we have a dataset which have two classes A and B. The percentage of class A is 40% and percentage of class B is 60%. Calculate the Gini index and entropy of the dataset.**

**n**

**Ans:- Gini Index or  $\text{gini}(D)$  or  $I_G = 1 - \sum_{j=1}^n P_j^2$  where.  $P_j$  is the relative frequency of class  $j$  in  $D$**

**$j=1$**

**$D = \text{Dataset}$**

Gini is given as  $Gini = 1 - ((P_1)^2 + (P_2)^2)$  therefore,

$$1 - ((40/100)^2 + (60/100)^2) = 0.48$$

c

$$\text{Entropy } I_H = - \sum_{j=1}^c P_j \log_2(P_j)$$

j=1

therefore :

$$-((40/100)\log_2(40/100) + (60/100)\log_2(60/100)) = 0.96$$

**10. What are the advantages of Random Forests over Decision Tree?**

**Ans:-**

- 1) Easy to interpret. Straightforward graphical visualization is intuitive makes it easy for anyone to understand**
- 2) Useful to identify variables most helpful in prediction. It splits based on the attribute significance.**
- 3) Handles both continuous and categorical targets attributes.**
- 4) Performs well on large datasets**
- 5) Not influenced by outliers and missing values to an extent.**
- 6) Non parametric method- does not make an assumption about the form of the mapping function. As they do not make any assumption about mapping function they are free to learn any functional form present in the data**

**7) Trees can handle qualitative input variables without the need to create dummy variables**

**11. What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling.**

**Ans:-** Machine learning algorithm just sees number — if there is a vast difference in the range say few ranging in thousands and few ranging in the tens, and it makes the underlying assumption that higher ranging numbers have superiority of some sort. So these more significant number starts playing a more decisive role while training the model.

The machine learning algorithm works on numbers and does not know what that number represents. A weight of 10 grams and a price of 10 dollars represents completely two different things — which is a no brainer for humans, but for a model as a feature, it treats both as same.

**Sklearn has following four scalers primarily**

- 1. Minmax scaler**
- 2. Robust scaler**
- 3. Standard Scaler**
- 4. Normalizer.**

**Minmax scaler :-** Minmax scaler should be the first choice for scaling. For each feature, each value is subtracted by the minimum value of the respective feature and then divide by the range of original maximum and minimum of the same feature. It has a default range between [0,1].

**RobustScaler:-** RobustScaler can be used when data has high outliers and we want to subside their effects. But unimportant outliers should be removed in the first place. RobustScaler subtracts the column's median and divides by the interquartile range.

**StandardScaler:-** StandardScaler rescales each column to have 0 mean and 1 Standard Deviation. It standardizes a feature by subtracting the mean and dividing by the standard deviation. If the original distribution is not normally distributed, it may distort the relative space among the features.

**12. Write down some advantages which scaling provides in optimization using gradient descent algorithm.**

**Ans:--** We can speed up gradient descent by scaling because  $\theta$  descends quickly on small ranges and slowly on large ranges, and oscillates inefficiently down to the optimum when the variables are very uneven.

We can use fixed learning rate during training without worrying about learning rate decay.

It has straight trajectory towards the minimum and it is guaranteed to converge in theory to the global minimum if the loss function is convex and to a local minimum if the loss function is not convex.

It has unbiased estimate of gradients. The more the examples, the lower the standard error.

Optimization algorithm that is not iterative and simply solves for one point

**13. In case of a highly imbalanced dataset for a classification problem, is accuracy a good metric to measure the performance of the model. If not, why?**

**Ans:--** In case of a highly imbalanced dataset for a classification problem, accuracy is not a good metric to measure the performance of the model.

Classification accuracy is a metric that summarizes the performance of a classification model as the number of correct predictions divided by the total number of predictions.

It is easy to calculate and intuitive to understand, making it the most common metric used for evaluating classifier models. This intuition breaks down when the distribution of examples to classes is severely skewed.

Intuitions developed by practitioners on balanced datasets, such as 99 percent representing a skillful model, can be incorrect and dangerously misleading on imbalanced classification predictive modeling problems.

**14. What is "f-score" metric? Write its mathematical formula.**

**Ans:-** The F measure (F1 score or F score) is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

$$\text{F-Measure} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

**15. What is the difference between fit(), transform() and fit\_transform()?**

**Ans:-** The fit method, when applied to the training dataset, learns the model parameters (for example, mean and standard deviation). We then need to apply the transform method on the training dataset to get the transformed (scaled) training dataset. We could also perform both of these steps in one step by applying fit\_transform on the training dataset.

In summary, fit performs the training, transform changes the data in the pipeline in order to pass it on to the next stage in the pipeline, and fit\_transform does both the fitting and the transforming in one possibly optimized step.

in other words

**fit()=** fit function that calculates the value of the parameter.

**-transform()** transform function applied value of the parameter of the actual data

**and give the normalized value.**

**- fit\_transform()** function performs both in the same step.