MACHINE LEARNING – WORKSHEET 4

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?

Ans - D) kernel will be changed to linear

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- C) In case of classification problem, the prediction is made by taking mode of the class labels predicted by the component trees.

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

Ans- A) Gradient Descent algorithm can diverge from the optimal solution.

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

Ans- C)both bias and variance increase

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

Train accuracy=0.95

Test accuracy=0.75

Which of the following is true regarding the model?

Ans- B) model is underfitting

subjective answer

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.

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

Ans- Random forests are a strong modeling technique and much more robust than a single decision tree. They aggregate many decision trees to limit overfitting as well as error due to bias and therefore yield useful results.

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

Ans- Scaling is required to rescale the data and it's used when we want features to be compared on the same scale for our algorithm. And, when all features are in the same scale, it also helps algorithms to understand the relative relationship better.

The comparative scales can further be divided into the following four types of scaling techniques: (a) Paired Comparison Scale, (b) Rank Order Scale, (c) Constant Sum Scale, and (d) Q-sort Scale.

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

Ans- Optimization refers to the task of minimizing/maximizing an objective function f(x) parameterized by x. In machine/deep learning terminology, it’s the task of minimizing the cost/loss function J(w) parameterized by the model’s parameters w ∈ R^d. Optimization algorithms (in case of minimization) have one of the following goals:

* Find the global minimum of the objective function. This is feasible if the objective function is convex, i.e. any local minimum is a global minimum.
* Find the lowest possible value of the objective function within its neighborhood. That’s usually the case if the objective function is not convex as the case in most deep learning problems.

Gradient Descent is the most common optimization algorithm in machine learning and deep learning. It is a first-order optimization algorithm. This means it only takes into account the first derivative when performing the updates on the parameters. On each iteration, we update the parameters in the opposite direction of the gradient of the objective function J(w) w.r.t the parameters where the gradient gives the direction of the steepest ascent. The size of the step we take on each iteration to reach the local minimum is determined by the learning rate α. Therefore, we follow the direction of the slope downhill until we reach a local minimum.

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- The class imbalance problem is a fundamental problem in machine learning, and it has received much attention [[8](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B8), [9](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B9), [10](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B10), [11](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B11), [12](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B12), [13](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B13), [14](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B14)]. This problem is known as partial classification [[15](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B15)], nugget discovery [[16](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B16)], classification problem with imbalanced datasets [[17](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B17)], or datasets with rare classes [[18](https://www.intechopen.com/books/recent-trends-in-computational-intelligence/classification-problem-in-imbalanced-datasets#B18)]. Considering the importance of this issue, a large amount of techniques have been developed trying to address this problem. These proposals can be divided into three groups which depend on how they deal with class imbalance. First, the algorithm-level approaches can either propose specific algorithms or modify the existing ones. Second, the data-level techniques introduce an additional processing step to decrease the effect of skewed class distribution such as undersampling and oversampling methods. Finally, the hybrid-level methods combine algorithm level and data level such as boosting and cost-sensitive learning.

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

Ans- The F-score, also called the F1-score, is a measure of a model’s accuracy on a dataset. It is used to evaluate binary classification systems, which [classify](https://deepai.org/machine-learning-glossary-and-terms/classifier) examples into ‘positive’ or ‘negative’.

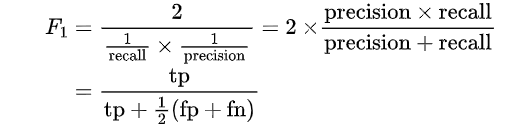
The F-score is a way of combining the [precision and recall](https://deepai.org/machine-learning-glossary-and-terms/precision-and-recall) of the model, and it is defined as the [harmonic mean](https://deepai.org/machine-learning-glossary-and-terms/harmonic-mean) of the model’s precision and recall.

The F-score is commonly used for evaluating information retrieval systems such as search engines, and also for many kinds of [machine learning](https://deepai.org/machine-learning-glossary-and-terms/machine-learning) models, in particular in [natural language processing](https://deepai.org/machine-learning-glossary-and-terms/natural-language-processing).

It is possible to adjust the F-score to give more importance to precision over recall, or vice-versa. Common adjusted F-scores are the F0.5-score and the F2-score, as well as the standard F1-score.

**F-score Formula**

The formula for the standard F1-score is the harmonic mean of the precision and recall. A perfect model has an F-score of 1.



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

Ans- The fit() function calculates the values of these parameters. The transform() function applies the values of the parameters on the actual data and gives the normalized value. The fit\_transform() function performs both(fit() , transform()) steps.