**MACHINE LEARNING**

ASSIGNMENT - 7

1. Which of the following in sk-learn library is used for hyper parameter tuning?

A) GridSearchCV() B) RandomizedCV()

C) K-fold Cross Validation D) All of the above

Answer: A

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

A) Random forest B) Adaboost

C) Gradient Boosting D) All of the above

Answer: A

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?

A) The regularization will increase B) The regularization will decrease

C) No effect on regularization D) kernel will be changed to linear

Answer:

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?

A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.

B) It denotes the number of children a node can have.

C) both A & B

D) None of the above

Answer: C

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

A) It's an ensemble of weak learners.

B) The component trees are trained in series

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

D)None of the above

Answer:

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

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

B) Gradient Descent algorithm can keep oscillating around the optimal solution and may not settle.

C) Both of them

D) None of them

Answer: D

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

A) Bias will increase, Variance decrease B) Bias will decrease, Variance increase

C)both bias and variance increase D) Both bias and variance decrease.

Answer: B

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

Train accuracy=0.95 and Test accuracy=0.75

Which of the following is true regarding the model?

A) model is underfitting B) model is overfitting

C) model is performing good D) None of the above

Answer:

**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.

Answer:

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

Answer: Random forests consist of multiple single trees each based on a random sample of the training data. They are typically more accurate than single decision trees.

Random forest is an ensemble method in which a classifier is constructed by combining several different *Independent*base classifiers. The independence is theoretically enforced by training each base classifier on a training set sampled with replacement from the original training set. This technique is known as bagging, or bootstrap aggregation. In Random Forest, further randomness is introduced by identifying the best split feature from a random subset of available features.

The ensemble classifier then aggregates the individual predictions to combine into a final prediction, based on a majority voting on the individual predictions.

It can be shown that an ensemble of independent classifiers, each with an error rate e, when combined significantly reduces the error rate.

Suppose we have 10 *independent*classifiers, each with error rate of 0.3

ϵ=0.3ϵ=0.3

In this setting, the error rate of the ensemble can be computed as below (we are taking a majority vote on the predictions. An ensemble makes a wrong prediction only when more than half of the base classifiers are wrong)

ϵensemble=∑i=10i=6(10i)ϵi(1−ϵ)10−i≈0.05ϵensemble=∑i=6i=10(10i)ϵi(1−ϵ)10−i≈0.05

It can be seen that with the theoretical guarantees stated above an ensemble model performs significantly well.

However in practice it is not possible to guarantee such classifier independence as they are trained from the same data, but still introduction of randomness helps achieve independence to a certain degree and it has been empirically observed that ensembles perform significantly well over individual base classifiers.

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

Answer: Scaling is essential for machine learning algorithms that calculate distances between data. Therefore, the range of all features should be normalized so that each feature contributes approximately proportionately to the final distance.

The most common techniques of feature scaling are Normalization and Standardization.

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

Answer: Batch Gradient Descent is when we sum up over all examples on each iteration when performing the updates to the parameters. Therefore, for each update, we have to sum over all examples:

https://miro.medium.com/max/60/1*nu5id-pd3BNCl1KktBxP4g.png?q=20

https://miro.medium.com/max/654/1*nu5id-pd3BNCl1KktBxP4g.png

for i in range(num\_epochs):  
 grad = compute\_gradient(data, params)  
 params = params — learning\_rate \* grad

The main advantages:

* 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.

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?

* Answer: Accuracy and error rate are the de facto standard metrics for summarizing the performance of classification models.
* Classification accuracy fails on classification problems with a skewed class distribution because of the intuitions developed by practitioners on datasets with an equal class distribution

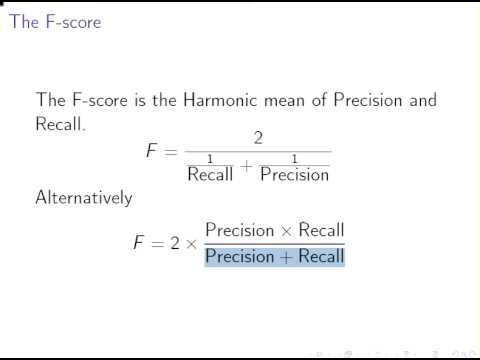
14. What is “f-score" metric? Write its mathematical formula.

Answer: 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.



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

Answer: The usage of methods fit(), transform(), fit\_transform() and predict() depend on the type of object.

**For Transformers:**

1. fit() - It is used for calculating the initial filling of parameters on the training data (like mean of the column values) and saves them as an internal objects state
2. transform() - Use the above calculated values and return modified training data
3. fit\_transform() - It joins above two steps. Internally, it just calls first fit() and then transform() on the same data.

**For Models:**

1. fit() - It calculates the parameters/weights on training data (e.g. parameters returned by coef() in case of Linear Regression) and saves them as an internal objects state.
2. predict() - Use the above calculated weights on test data to make the predictions
3. transform() - Cannot be used
4. fit\_transform() - Cannot be used