Practice Test - Practical Machine Learning

Total points 14/20

Duration - 30 Mins Total Marks : 20 Marks

Instructions :

- There are total 16 questions, all questions are compulsory.
- Que. no. 7 has 5 marks
- Select the option by clicking on checkbox shown thus ◯.

Email *

swalinimore@gmail.com

PRN No. *

220340128047

Name *

Swalini Mary M

✗ 1. As an input for calculating log loss for model evaluation of classification algorithm, we require *0/1

- a. Response variable values of test set(y_test) and predicted probabilities from the algorithm on test data (output from predict_proba(X_test))
- b. Response variable values of train set(y_train) and predicted probabilities from the algorithm on test data(output from predict_proba(X_test))
- c. Response variable values of test set(y_test) and predictions from the algorithm on test data (output from predict(X_test)) ✗
- d. Response variable values of test set(y_test) and predictions from the algorithm on train data(output from predict(X_train))

Correct answer

- a. Response variable values of test set(y_test) and predicted probabilities from the algorithm on test data (output from predict_proba(X_test))

✗ 2. In which (unsupervised) clustering method is clustering of observations based on the nearness (neighbourhood) of the points *0/1

- a. K-Means ✗
- b. Hierarchical
- c. DBSCAN
- d. K-NN

Correct answer

- c. DBSCAN

✓ 3. Min Max scaler transforms the data to the range * 1/1

- a. (1,10)
- b. (-infinity, infinity)
- c. (0,infinity)
- d. (0,1) ✓

✗ 4. There is an option in Python to calculate area under the curve of ROC Curve in case of Multi-Class problem. *0/1

- a. TRUE
- b. FALSE ✗

Correct answer

- a. TRUE

✗ 5. What is different with the trees grown with random forest? * 0/1

- a. Nothing different same as single decision tree ✗
- b. The best split is decided by cost function
- c. The best split is decided only on a small number of randomly selected features
- d. The best split is decided only on a small number of randomly selected observations

Correct answer

- c. The best split is decided only on a small number of randomly selected features

✗ 6. The principal component analysis transformation ultimately gives us *0/1

- (●) a. All the observations which are independent of each other ✗
- ( ) b. All the features which are independent of each other
- ( ) c. All the observations which are dependent on each other
- ( ) d. All the features which are dependent on each other

Correct answer

- (●) b. All the features which are independent of each other

✓ 7. Match the following (5 marks) * 5/5

Model	***Important Hyper-parameters***
1. support vector machines - radial	a. k (n_neighbors)
2. decision trees	b. min_samples_split, min_samples_leaf, max_depth
3. k-nearest neighbours	c. learning_rate , max_depth, n_estimators
4. random forest	d. max_features
5. gradient boosting	e. C, gamma

- ( ) a. 1-a,2-b,3-c,4-d,5-e
- ( ) b. 1-e,2-d,3-a,4-a,5-c
- (●) c. 1-e,2-b,3-a,4-d,5-c ✓
- ( ) d. 1-d,2-b,3-a,4-e,5-c

✓ 8. Which of the statements is true with parameters and hyper-parameters? *1/1

- (●) a. Hyper-parameters are decided before the program is executed, whereas parameters are calculated from the data ✓
- ( ) b. Parameters are decided before the program is executed, whereas hyper-parameters are calculated from the data
- ( ) c. Parameters and Hyper-parameters are same
- ( ) d. None of Above

✓ 9. Support Vector Machines in Python cannot be performed with Multi-Class classification problem *1/1

- ( ) a. TRUE
- (●) b. FALSE ✓

✓ 10. In the cost function J( ), of decision tree algorithm, which of the metric is not included? *1/1

- (●) a. Area Under the Curve Score of ROC ✓
- ( ) b. Gini's Index
- ( ) c. Squared Error
- ( ) d. Log Loss

✓ 11. Concept of Simple random sampling with replacement is used in which of the ML algorithms? *1/1

- a. Support Vector Machines
- b. Linear Regression
- c. Bagging ✓
- d. Stacking

✓ 12. The argument (random_state=...) is specified * 1/1

- a. So that every time the function is run, the output gets different
- b. So that every time the function is run, the output gets same ✓
- c. Because it is compulsory argument
- d. None of these

✓ 13. Log loss in Python can be computed only for binary classification and not for multi-class classification *1/1

- a. TRUE
- b. FALSE ✓

✓ 14. Which of the following is not evaluation metrics for classification algorithms? *1/1

- (●) a. R2 score ✓
- ( ) b. Area under ROC
- ( ) c. F1 Score
- ( ) d. Precision

✗ 15. The function in scikit-learn named GridSearchCV.fit( ) always finds the best parameter set *0/1

- ( ) a. based on maximum of the score mentioned in its argument option “scoring=”
- ( ) b. based on minimum of the score mentioned in its argument option “scoring=”
- ( ) c. depends upon ‘random_state=’ option
- (●) d. None of Above ✗

Correct answer

- (●) a. based on maximum of the score mentioned in its argument option “scoring=”

✓ 16. K-Means clustering doesn’t start with randomization. * 1/1

- ( ) a. TRUE
- (●) b. FALSE ✓