MACHINE LEARNING ASSIGNMENT AND PYTHON WORKSHEET 1

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

A) Least Square Error B) Maximum Likelihood

C) Logarithmic Loss D) Both A and B

ANSWER = least square

2) Which of the following statement is true about outliers in linear regression?

A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers

C) Can't say D) none of these

ANSWER= Linear regression is sensitive to outliers

3) A line falls from left to right if a slope is _____?

A) Positive B) Negative C) Zero D) Undefined

ANSWER = Negative

4) Which of the following will have symmetric relation between dependent variable and independent variable?

A) Regression B) Correlation C) Both of them D) None of these

ANSWER = Both of them

5) Which of the following is the reason for over fitting condition?

A) High bias and high variance B) Low bias and low variance

C) Low bias and high variance D) none of these

ANSWER = Low bias and high variance

6) If output involves label then that model is called as:

A) Descriptive model B) Predictive modal

C) Reinforcement learning D) All of the above

ANSWER = Predictive model

7) Lasso and ridge regression techniqes belong to_____?

A) cross validation B) removing outlier

C) smote D) regularization

ANSWER = Regularization

8) To overcome with imbalance dataset which technique can be used?

A) Cross validation B) Regularization

C) Kernel D) SMOTE

ANSWER = Smote

9) The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

A) TPR and FPR B) Sensitivity and precision

C) Sensitivity and Specificity D) Recall and precision

ANSWER = TPR AND FPR

10) In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

A) True B) False

ANSWER = FALSE

11) Pick the feature extraction from below:

A) Construction bag of words from a email

B) Apply PCA to project high dimensional data

C) Removing stop words

D) Forward selection

ANSWER = Apply Pca to project high dimensional data

12) Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

A) We don’t have to choose the learning rate.

B) It becomes slow when number of features is very large.
C) We need to iterate.
D)It does not make use of dependent variable

ANSWER = It does not make use of dependent variable

3

Q (13) EXPLAIN THE TERM OF REGULARIZATION ?

ANSWER = Regularization is a technique used in machine learning to prevent overfitting and improve the generalization of models. It involves adding a penalty term to the loss function during training to discourage the model from fitting the noise in the training data. Regularization can be explicit, where a penalty term is added to the optimization function, or implicit, where techniques like early stopping or using a robust loss function are employed. There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization. Ridge regularization modifies overfitted or underfitted models by adding a penalty equivalent to the sum of the squared weights, while Lasso regularization adds a penalty equivalent to the sum of the absolute values of the weights. Regularization is an important technique in machine learning that helps to improve the performance of models on unseen data.

Q (14) Which particular algorithms are used for regularization?

ANSWER = Regularization is a technique used in machine learning to prevent overfitting and improve the generalization of a model by adding a penalty term to the cost function. Various algorithms employ regularization methods to achieve this. Here are some commonly used regularization algorithms:

1. **L1 Regularization (Lasso):** This algorithm adds the absolute values of the coefficients as a penalty term to the cost function, encouraging sparsity in the model.
2. **L2 Regularization (Ridge):** Ridge regression adds the squared values of the coefficients as a penalty term, helping to prevent overly large coefficients.
3. **Elastic Net:** This algorithm combines both L1 and L2 regularization, providing a balance between feature selection and coefficient shrinkage.
4. **Dropout:** Commonly used in neural networks, dropout randomly deactivates a fraction of neurons during training, preventing reliance on specific neurons and promoting a more robust model.
5. **Weight Decay:** It is a general term for regularization techniques that penalize large weights. L1 and L2 regularization are specific instances of weight decay.

6. **Early Stopping:** This is a technique where training stops once the performance on a validation set starts to degrade, preventing the model from overfitting the training data.
7. **Batch Normalization:** While primarily used for normalization, it can act as a regularizer by adding a slight noise during training, which can improve generalization.
8. **Droupout2D (for Convolutional Neural Networks):** Similar to dropout but applied to entire feature maps in convolutional layers.
9. **Max Norm Constraints:** Limits the maximum value of the weights during training, preventing them from growing too large.
10. **Orthogonalization:** Ensures that the weight vectors are orthogonal, adding a constraint that prevents the model from fitting noise.

These regularization techniques play crucial roles in enhancing the performance and robustness of machine learning models by discouraging them from fitting noise and capturing unnecessary complexity in the training data.

Q (15) Explain the term error present in linear regression equation ?

ANSWER The error term in a linear regression equation represents the difference between the observed data and the true population data. It accounts for the effect of variables that are not included in the model. This term is often denoted by ε and encompasses factors such as omitted variables, measurement errors, and unpredictable effects. While the error term is a population value that is unobservable, the residuals, which are the sample estimates of the error for each observation, are used in practice. Violations of the assumptions related to the error term can lead to biased estimates in the regression model. The error term is a fundamental concept in regression analysis, and it is essential to consider its properties and implications when interpreting the results of a linear regression model.

The error term in a linear regression equation represents the difference between the observed data and the true population data. It accounts for the effect of variables that are not included in the model. This term is often denoted by ε and encompasses factors such as omitted variables, measurement errors, and unpredictable effects. While the error term is a population value that is unobservable, the residuals, which are the sample estimates of the error for each observation, are used in practice. Violations of the assumptions related to the error term can lead to biased estimates in the regression model. The error term is a fundamental concept in regression analysis, and it is essential to consider its properties and implications when interpreting the results of a linear regression model

1. QUESTION(1) Which of the following operators is used to calculate remainder in a division?

A) # B) &

C) % D) $

ANSEWER= C %

2) In python 2//3 is equal to?

A) 0.666 B) 0 C) 1 D) 0.67

ANSWER = C 1

3) In python, 6<<2 is equal to?

A)36 B) 10

C) 24 D) 45

ANSWER = 24

4) In python, 6&2 will give which of the following as output?

A) 2 B) True C) False D) 0

ANSWER= 2

5) In python, 6|2 will give which of the following as output?

A) 2 B) 4

C)0 D) 6

ANSWER = 6

6) What does the finally keyword denotes in python?

A) Itis used to mark the end of the code

B)It encloses the lines of code which will be executed if any error occurs while executing the lines of code in the try block.

c) the finally block will be executed no matter if the try block raises an error or not.
d)None of the above

ANSWER = The Finally block will be excuted no matter if the try block reises an error or not.

7) What does raise keyword is used for in python?
a) It is used to raise an exception. B) It is used to define lambda function
C) it's not a keyword in python. D) None of the above

ANSWER = It is used to raise an exception

8) Which of the following is a common use case of yield keyword in python?
A)in defining an iterator B) while defining a lambda function
C) in defining a generator D) in for loop.

ANSWER = In defining a generator

9) Which of the following are the valid variable names?
A)__abc B) 1abc
C) abc2 D) None of the above

ANSWER= abc2

10) Which of the following are the keywords in python?
A) yield B) raise
C) look-in D) all of the above

ANSWER = all of the above