

MACHINE-LEARNING-WORKSHEET-1

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

Answer: A) Least Square Error

B) Maximum Likelihood

C) Logarithmic Loss

D) Both A and B

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

Answer: A) Linear regression is sensitive to outliers

B) linear regression is not sensitive to outliers

C) Can't say

D) none of these

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

A) Positive

Answer: B) Negative

C) Zero

D) Undefined

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

A) Regression

Answer: B) Correlation

C) Both of them

D) None of these

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

A) High bias and high variance

B) Low bias and low variance

Answer: C) Low bias and high variance

D) none of these

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

A) Descriptive model

Answer: B) Predictive model

C) Reinforcement learning

D) All of the above

7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation

B) Removing outliers

C) SMOTE

Answer: D) Regularization

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

A) Cross validation

B) Regularization

C) Kernel

Answer: D) SMOTE

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

Answer: A) TPR and FPR

B) Sensitivity and precision

C) Sensitivity and Specificity

D) Recall and precision

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

A) True

Answer: B) False

11. Pick the feature extraction from below:

A) Construction bag of words from a email

Answer: B) Apply PCA to project high dimensional data

C) Removing stop words

D) Forward selection

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

Answer: A) We don't have to choose the learning rate.

Answer: 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.

13. Explain the term regularization?

Answer: Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

14. Which particular algorithms are used for regularization?

Answer: There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization.

Ridge Regularization:

Also known as Ridge Regression, it modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

Lasso Regression:

It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients

15. Explain the term error present in linear regression equation?

Answer: An error term in statistics is a value which represents how observed data differs from actual population data. It can also be a variable which represents how a given statistical model differs from reality. The error term is often written ϵ .

Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

1. measuring the distance of the observed y-values from the predicted y-values at each value of x;
2. squaring each of these distances;
3. calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.