

MACHINE LEARNING ASSIGNMENT

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

Answer. Least Square Error.

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

Answer. Linear regression is sensitive to outliers.

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

Answer Negative.

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

Answer. The relation is symmetric in correlation but in regression it will not therefore answer is correlation.

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

Answer. SMOTE

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

Answer. Predictive model

7. Lasso and Ridge regression techniques belong to _____

Answer. Cross validation

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

Answer. Kernel

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

Answer- TPR and FPR

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

Answer- True

11. Pick the feature extraction from below:

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?

Answer- We don't have to choose the learning rate.

It becomes slow when number of features is very large .

We need to iterate.

13. Explain the term regularization?

Answer- Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoiding overfitting.

The commonly used regularization techniques are -

1. L1 regularization
2. L2 regularization
3. Dropout regularization

14. Which particular algorithms are used for regularization?

Answer- There are three types of regularization algorithm -

1. Ridge Regression
2. LASSO (Least Absolute Shrinkage and Selection Operator) Regression
3. Elastic -Net Regression.

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

Answer- The error term is the stuff is not explain by the model.

For a very simple example, suppose you are predicting the weight of adult human males based on their height. Well height is certainly related to weight-taller tend to be heavier- but the model won't be perfect because there is a range of weights and each weights. The error is the difference between the predicted value and the actual value.

PYTHON WORKSHEET

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

Answer- %

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

Answer- 0

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

Answer- 24

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

Answer- 2

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

Answer- 6

6. What does the finally keyword denotes in python?

Answer- It encloses the lines of code which will be executed if any error occurs while executing the lines of code in the try block

7. What does raise keyword is used for in python?

Answer- It is used to raise an exception.

8. Which of the following is a common use case of yield keyword in python?

Answer- in defining a generator

9. Which of the following are the valid variable names?

Answer- _abc , abc2

10. Which of the following are the keywords in python?

Answer- yield and raise

STATISTICS WORKSHEET

1. Bernoulli random variables take (only) the values 1 and 0

Answer- True

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

Answer- Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?

Answer- Modeling bounded count data

4. Point out the correct statement.

Answer- All of the mentioned

5. _____ random variables are used to model rates.

Answer- Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

Answer- False

7. 1. Which of the following testing is concerned with making decisions using data?

Answer- Hypothesis

8. 4. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

Answer- 0

9. Which of the following statement is incorrect with respect to outliers?

Answer- None of the mentioned

10. What do you understand by the term Normal Distribution?

Answer- Normal distribution, also known as the Gaussian distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph normal distribution appear as bell curve.

11. How do you handle missing data? What imputation techniques do you recommend

Answer- Missing data is an inevitable part of the process. Lots of data resources, time and energy into making sure the data is as accurate as possible. But somehow data goes missing.

The imputation techniques which is recommended are-

1. Deletion methods
2. Regression Analysis
3. Substitute the value such as means

12. What is A/B testing?

Answer- A/B testing is a user experience research methodology. A/B tests consist of a randomised experiment with two variants, A and B . It includes application of statical hypothesis testing or two-sample hypothesis testing as used in the field of statistics.

13. Is mean imputation of missing data acceptable practice?

Answer- No, it is typically considered terrible practice since it ignores feature correlation.

14. What is linear regression in statistics?

Answer- In statistics, linear regression is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression and for more than one its multiple linear regreesion.

15. What are the various branches of statistics?

Answer- There are three branches of statistics –

1. Data Collection
2. Descriptive Statistics
3. Inferential Statistics