FLIP ROBO

Assignment No.1

MACHINE LEARNING

NANE   ARCHIT SINGHAL

INTERENSHIP 31

Q No.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

Q No.2   Which of the following the statement is true about outliers in Linear Regression?

Answer: A) Linear regression is sensitive to outliers

Q No.3   A line falls from left to right if a slope is_?

Answer: B) Negative

Q No.4   Which of the following will have symmetric relation between dependent variable and independent variable?

Answer: D) None of these

Q No.5   Which of the following reason for over fitting condition?

Answer: D) None of these

Q No.6 If output involves label then that model is called as:

Answer: B) Predictive modal

Q No.7   Lasso and Ridge Regression techniques belong to_.

Answer: D) Regularization

Q No.8   To overcome with imbalance dataset which techniques can be used?

Answer: D) SMOTE

Q No.9   The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Answer: C) Sensitivity and Specificity

Q No.10   In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

Answer: A) TRUE

Q No.11 Pick the feature extraction from below:

Answer: B) Apply PCA to project high dimensional data.

Q No.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.

           B) It becomes slow when number of features is very large.

Q No.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.

Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

Regularization   Techniques

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

Q No. 14   Which particular algorithms are used for regularization?

Answer: Ridge Regression

Ridge regression is also called L2 norm or regularization.

When using this technique, we add the sum of weight's square to a loss function and thus create a new loss function which is denoted thus:

$$\text{Loss} = \sum_{j=1}^{m}\left(Yi - Wo - \sum_{i=1}^{n} Wi\, Xji\right)^{2} + \lambda \sum_{i=1}^{n} Wi^{2}$$

As seen above, the original loss function is modified by adding normalized weights. Here normalized weights are in the form of squares.

You may have noticed parameters λ along with normalized weights. λ is the parameter that needs to be tuned using a cross-validation dataset. When you use $\lambda=0$, it returns the residual sum of square as loss function which you chose initially. For a very high value of λ, loss will ignore core loss function and minimize weight's square and will end up taking the parameters' value as zero.

Now the parameters are learned using a modified loss function. To minimize the above function, parameters need to be as small as possible. Thus, L2 norm prevents weights from rising too high.

Q No.15   Explain the term error present in linear regression equation?

Answer :   I do not answer of this question .

FLIP ROBO

Worksheet

PYTHON WORKSHEET 1

NAME ARCHIT SINGHAL

INTERENSHIP 31

Q No.1 Which of the following operators is used to calculate remainder in a division?

Answer: C) %

Q No.2 In python 2//3 is equal to?

Answer: B) 0

Q No.3 In python, 6<< is equal to?

Answer: C) 24

Q No.4 In python, 6&2 will give which of the following as output?

Answer: A) 2

Q No.5 In python, 6|2 will give which of the following as output?

Answer: D) 6

Q No.6 What does the finally keyword denotes in python?

 Answer: D) None of the above

Q No.7 What does raise keyword is used for in python?

Answer: A) It is used to raise an exception

Q No.8 Which of the following is a common use case of yield keyword in python?

Answer: A) in defining an iterator

Q No.9 Which of the following are the valid variable names?

Answer: A) _abc

       C) abc2

Q No.10 Which of the following are the keywords in python?

Answer: A) Yield

       B) Raise

FLIP ROBO

Worksheet

STATISTICS WORKSHEET-1

Name Archit Singhal

Interenship 31

Q No.1 Bernoulli random variables take (only) the values 1 and 0.

Answer: a) True

Q No.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: a) Central Limit Theorem

Q No.3 Which of the following is incorrect with respect to use of Poisson distribution?

Answer: b) Modeling bounded count data

Q No.4 Point out the correct statement.

Answer: d) All of the mentioned

Q No.5 _____ random variables are used to model rates.

Answer: c) Poisson

Q No.6 10. Usually replacing the standard error by its estimated value does change the CLT.

Answer: a) True

Q No.7 1. Which of the following testing is concerned with making decisions using data?

Answer: b) Hypothesis

Q No.8 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.

Answer: a) 0

Q No.9 Which of the following statement is incorrect with respect to outliers?

Answer: d) None of the mentioned

Q No.10 What do you understand by the term Normal Distribution?

Answer: Normal distribution, also known as the Gaussian distribution, is a probability 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 graphical form, the normal distribution appears as a "bell curve".

Q No.11 How do you handle missing data? What imputation techniques do you recommend?

Answer: Overview of Missing Data
1. Deletions. Pairwise Deletion. Listwise Deletion/ Dropping rows. Dropping complete columns.
2. Basic Imputation Techniques. Imputation with a constant value. Imputation using the statistics (mean, median, mode)
3. K-Nearest Neighbor Imputation

There are the following some imputation techniques:

* Mean imputation. Simply calculate the mean of the observed values for that variable for all individuals who are non-missing. ...
* Substitution. ...
* Hot deck imputation. ...
* Cold deck imputation. ...
* Regression imputation. ...
* Stochastic regression imputation. ...
* Interpolation and extrapolation.

Q No.12 What is A/B testing?

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

Q No.13 Is mean imputation of missing data acceptable practice?

Answer: Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

Q No.14 What is linear regression in statistics?

Answer: Linear regression analysis is **used to predict the value of a variable based on the value of another variable**. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Q No.15 What are the various branches of statistics?

Answer: There are three types branches of statistics:

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