DA BodhiTree MCQs

Which type of test is the Wilcoxon rank sum test?

Answer

non-Parametric

Input data for Wilcoxon test is normally distributed, True or False?

Answer

False

What is the null hypothesis for a Wilcoxon test

Answer

Two group means are equal.

Which of following test statics is used in Wilcoxon Rank Sum Test?

Answer

test statistics <= critical value, Ho will be Rejected

What must you include when applying Wilcoxon Rank sum test?

Answer

“Critical Value”, “Rank sum”

Type 1 error is also called as

Answer

False Positive

Type 2 error is also called as

Answer

False negative

Type 1 error occurs when_____

Answer

Null hypothesis rejected when it is true.

Type 2 error occurs when_____

Answer

Null hypothesis is accepted when it is false

How to reduce Type 2 error?

Answer

By increasing sample size

Analysis of Variance is statistical method of comparing____of several populations

Answer

Means

ANOVA is used when____

Answer

If more than two population

What is Null Hypothesis in ANOVA?

Answer

all group means are equal

What does ANOVA calculate?

Answer

F ratio

What are the two types of variance which can occur in your data?

Answer

Between and within groups

If between group mean sum of square variability increases value of F statistics_____

Answer

Increases

What must you include when applying ANOVA test?

Answer

“Means”, “Critical Value”, “degree of freedom”, “F statistics”

How many dependent variables are there in a two-way ANOVA?

Answer

2

Which of following test statics is used in ANOVA?

Answer

if critical value < F ratio, Ho will be Rejected

Various types of ANOVA are___

Answer

“Two way ANOVA”, “ANCOVA”, “MANOVA”

Clustering is the classification of objects into different groups, True or False?

Answer

True

Clustering partition the data into k subsets, True or False

Answer

True

Clustering extracts the known patterns from the existing data, True or False?

Answer

False

Clustering techniques falls in the category of_____

Answer

Unsupervised learning

K Means falls in the category of

Answer

Partitional Clustering

Agglomerative clustering is an ____ approach

Answer

bottom-up

Which of the following is determined by Distance Measure?

Answer

Similarity of two elements

K Means is _____

Answer

Centroid based method

Clustering is often used as a lead-in to classification, True or False?

Answer

True

K Means algorithm is iterative in nature, True or False?

Answer

True

In K Means algorithm, K is_____

Answer

Number of centers

WSS metric is the sum of the squares of the distances between each data point and the_____.

Answer

closest centroid

To use k-means properly, it is important to

Answer

All of the above

Once the clusters are identified, it is often useful to label them in a descriptive way.True or False?

Answer

True

K-means is easily applied to ……

Answer

Numeric attributes

Identification of highly correlated attributes is important for __.

Answer

reduction

K-means sensitive to starting seeds, True or False?

Answer

True

The process of identifying the appropriate value of k is referred to as finding the_____.

Answer

elbow

A _____ is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility

Answer

Decision tree

Decision Trees can be used for Classification Tasks.

Answer

True

Tree/Rule based classification algorithms generate … rule to perform the classification.

Answer

if-then

What is Decision Tree?

Answer

Flow-Chart & Structure in which internal node represents test on an attribute, each branch represents outcome of test and each leaf node represents class label

What is the approach of basic algorithm for decision tree induction?

Answer

Greedy

High entropy means that the partitions in classification are

Answer

not pure

What is information gain of attribute?

Answer

It is a measure of purity

Gini index is used by_____.

Answer

CART

How will you counter over-fitting in decision tree?

Answer

By pruning the longer rules

Calculate entropy of following equation?

$$I(all\_data) = -\left[\left(\frac{9}{16}\right)\log_2\left(\frac{9}{16}\right)+\left(\frac{7}{16}\right)\log_2\left(\frac{7}{16}\right)\right]$$

Answer

0.9836

What is true about Data Visualization?

Answer

All of the above

Which one of the following is most basic and commonly used techniques?

Answer

“Line charts”

Common use cases for data visualization include?

Answer

“All of the above”

Which of the follwoing are Data Visulaization tool?

Answer

“All of above”

Which of the following is false?

Answer

“Data visualization decrease the insights and take solwer decisions”

Which among the following are the features of the Hadoop?

Answer

All of Above

Hadoop Framework is written in_____.

Answer

Java

Which of the following is component of Hadoop?

Answer

All of Above

As compared to RDBMS, Hadoop_____

Answer

works better on unstructured and semistructured data.

Which of the following is the daemon of Hadoop?

Answer

All of Above

Which type of data Hadoop can deal with is?

Answer

All of Above

When a client contacts the namenode for accessing a file, the namenode responds with____

Answer

Block Id and hostname of all the data nodes containing that block.

Which of the following is used for machine learning on Hadoop?

Answer

Mhoot

Zookeeper ensures that_____.

Answer

Only one namenode is actively serving the client requests are

The tables created in hive are stored as____.

Answer

a subdirectory under the database directory