R / Machine Learning

Email address *

junaidkh190@gmail.com

When numerical values in a dataset have different measurement scales in kNN, it is best to *

- ○ Use datatypes that are similar
- ○ Remove columns that are not similar
- ◉ Standardize the dataset
- ○ Resample

Logistic regression can be used for *

- ○ Classification only
- ◉ Regression only
- ○ Both Classification and Regression
- ○ Clustering

To identify the best selling product, one can use the function *

- ( ) sort()
- (•) rank()
- ( ) order()
- ( ) trim()

Goal of Linear Regression is *

- ( ) Maximise SSE
- (•) Minimize SST
- ( ) Minimise SSE
- ( ) Minimize SSR

Machine Learning is possible due to *

- ( ) More training data and less test data
- ( ) Processed data
- ( ) Generalization
- (•) Selection of right algorithm

Eliminating x-variables from a dataset during Model building process is a *

- (●) Feature selection process
- ( ) Data cleansing process
- ( ) Data enhancing process
- ( ) Bad choice

To concatenate 2 dataframes having the same number of columns, we use *

- ( ) df1+df2
- ( ) df1-df2
- (●) cbind()
- ( ) rbind()

Misclassified data represented in the ConfusionMatrix are *

- (●) FP,FN
- ( ) TP,TN
- ( ) FN,TP
- ( ) FP,TN

Function to join multiple strings into a single string is *

- ○ substr()
- ○ join()
- ○ print()
- ◉ paste()

A vector contains 12 marks of 4 students; each student having 3 subjects (marks represented in a sequence for each subject).How would this be represented in a matrix ? *

- ◉ matrix(<data>,ncol=3,byrow=T)
- ○ matrix(<data>,ncol=4,byrow=T)
- ○ matrix(<data>,nrow=3,byrow=F)
- ○ matrix(<data>,nrow=4,byrow=F)

Error terms in a logistic regression has a normal distribution. This statement is *

- ○ An assumption
- ○ A necessary condition
- ◉ May be true sometimes
- ○ False

To select all records from a dataframe (stud) where column "email" id has the value "yahoo" *

- ○ stud[grep("Yahoo", email),]
- ◉ stud[grep(stud$email, "Yahoo"),]
- ○ stud[grep("yahoo", stud$email),]
- ○ stud[grep(email, "yahoo"),]

Random creation of salaries of 100 people is an example of *

- ○ Selection with replacement
- ◉ Selection without replacement
- ○ Biased Selection
- ○ Normal distribution

kNN works on the basis of *

- ○ Similarity functions
- ○ Dissimilarity functions
- ◉ Distance functions
- ○ Linear functions

Relationships caused by Linear regression is not always *

- ○ Linear
- ○ Independent
- ○ Causal
- ◉ Good

Right sequence of a model building process *

- ○ Read data, Predict, Build Model, Summarise Model, Identify significant parameters
- ○ Read data, Identify significant parameters, Summarise Model, Build Model, Predict
- ◉ Read data, Build Model, Summarise Model, Identify significant parameters, Predict
- ○ Read data, Build Model, Identify significant parameters, Predict, Summarise Model

mylist = list("type1"=list("a", "b"), "type2"=list("c","d")) is a *

- ○ Named single list
- ○ Named nested list
- ○ Unnamed single list
- ◉ Unnamed nested list

month.abb is a *

- ( ) Variable name
- (•) Pre-defined list
- ( ) Constant vector
- ( ) R-dataframe

seq(<n>,<n1>,<n2>) will generate *

- ( ) n numbers between <n1> and <n2>
- ( ) <n1> numbers between <n> and <n2>
- ( ) random sequence of <n> numbers between <n1> and <n2>
- (•) sequence of numners between <n1> and <n2> with an interval of n

When a factor variable in a dataset is blank, the easiest way to impute that data is by using *

- (•) Median
- ( ) Standard Deviation
- ( ) Range
- ( ) Mode