PYTHON-WORKSHEET 1 ANSWERS

Q1) Option C-%

Q2) Option B-0

Q3) Option C-24

Q4) Option A-2

Q5) Option D-6

Q6) Option C- the finally block will be executed no matter if the try block raises an error or not.

Q7) Option A- It is used to raise an exception.

Q8) Option C- in defining a generator

Q9) Option A & Option C

Q10) Option A & Option B

Q11 to Q15 are in **python jupyter notebook file**.

STATISTICS WORKSHEET-1

Q1) Option b)- False

Q2) Option a)- Central Limit Theorem

Q3) Option b)-Modelling bounded count data

Q4) Option d)-All of the mentioned

Q5) Option c)-Poisson

Q6) Option b)- False

Q7) Option b)-Hypothesis

Q8) Option a)-0

Q9) Option c)- Outliers cannot conform to the regression relationship

Q10) What do you understand by the term Normal Distribution?

ANSWER -Normal distribution is also known as Gaussian distribution, in which random variable values are plotted with respect to its frequency. The shape of this distribution is like a bell. In this distribution the mean, mode and median are same and at the centre of the plot i.e. at 0. It's a symmetric distribution with half values at the right and other half at the left. This distribution also represents probability distribution of random variable values.

Q11) How do you handle missing data? What imputation techniques do you recommend?

ANSWER- We handle missing data by Imputation. There are 3 techniques used for imputation:

1) Simple Imputer- In this technique simply average value of entire column is added to the null value.
2) Knn Imputer- In this technique the null values are imputed based on their relation with other columns.
3) Iterative Imputer- In this technique the columns with null values are treated as a label and other columns are treated as a feature then a model is build and the null values are then predicted, this is similar to linear regression.

Q12) What is A/B Testing?

ANSWER- A/B Testing also known as split testing is a method of testing user experiences related to any product. Let's understand this with the help of an example, suppose we have 100 customers and we have to observe these customers response for 2 versions of a product, version A and version B, so for this we will divide 100 customers into 2 groups of 50 customers each. To each group we will give version A and version B of the product and record their response.

The responses will further let the company decide which version is preferred and any other changes as well.

Q13) Is mean imputation of missing data acceptable practice?

ANSWER- Its easy practice to impute missing values by the mean of that feature but it has one big drawback that is if we impute the missing values using its mean value then we are doing bad to our data instead of good as this leads to massive changes in the correlation of features.

Q14) What is linear regression in statistics?

ANSWER- Linear Regression is an analysis used to predict the value of dependent variables based on some independent variables. The dependent variables are known as labels and independent variables are known as features. There are two types of linear regression:

1) Simple linear regression- In this dependent variable value is predicted using only single feature i.e., independent variable. The mathematical formula for this is:

   $$Y=mx+c$$

   Where Y is the predicted value,

   m is the coefficient or the slope of the linear line,

   c is the coefficient or y axis intercept.

2) Multiple Linear regression- In this dependent variable value is predicted using multiple features.

Q15) What are the various branches of statistics?

ANSWER- There are 2 branches of statistics:

1) Descriptive Statistics- This branch of statistics deals with describing the data either with the help of graphs like bar graph, histogram, pie chart or describing the data numerically by calculate mean, mode, median, variance, standard deviation etc.

2) Inferential Statistics- This branch of statistics deals with finding out conclusion from a given set of data. In Inferential statistics we get small samples from huge population of data and

we conclude or find out inference from that data sample and later that inference is induced for entire population.

MACHINE LEARNING WORKSHEET 1

Q1) Option A)- Least Square Error

Q2) Option B)- Linear regression is sensitive to outliers

Q3) Option B)-Negative

Q4) Option B)-Correlation

Q5) Option C)- Low bias and high variance

Q6) Option B)-Predictive modal

Q7) Option A)-Cross Validation

Q8) Option D)-SMOTE

Q9) Option C)- Sensitivity and Specificity

Q10) Option B)-False

Q11) Option B)- Apply PCA to project high dimensional data

Q12) Option A, B & C

Q13) Explain the term regularization?

ANSWER- Regularization is an important part of machine learning. This technique is used to avoid overfitting of model by optimizing the weights of the features thus it reduces error in a model and increases its accuracy.

Q14) Which particular algorithms are used for regularization?

ANSWER- There 2 algorithms that are used for regularization. They are:

1) Ridge (L2 norm)- This algorithm gives very less or no importance to the feature that has no contribution in predicting the label and the features that are highly contributing to those features it adds a penalty and their weights will be reduced,
2) LASSO (L1 norm)- this algorithm removes or vanishes those features those who doesn't contribute anything in the prediction of label and for highly contributing features penalty is added and their weights are reduced.

Q15) Explain the term error present in linear regression equation?

ANSWER- Error or residual is basically the difference between the actual target value and the predicted target value.

Error=predicted-actual

Error(r)=y-(mx+c)

Were,

Y is predicted value and (mx+c) is the actual value.

So linear regression gets the best fit line in such a way that the error or residual is minimized.