Python – Worksheet 1

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

Answer: **C) %**

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

Answer: **B) 0**

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

Answer: **C) 24**

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

Answer: **A) 2**

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

Answer: **D) 6**

1. What does the finally keyword denotes in python?

Answer: **C) The finally block will be executed no matter if the try block raises an error or not.**

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

Answer: **A) It is used to raise an exception**

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

Answer: **C) in defining a generator**

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

Answer: **C) abc2**

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

Answer: **D) all of the above**

1. Write a python program to find the factorial of a number.
2. Write a python program to find whether a number is prime or composite.
3. Write a python program to check whether a given string is palindrome or not.
4. Write a Python program to get the third side of right-angled triangle from two given sides.
5. Write a python program to print the frequency of each of the characters present in a given string.

**Answers of 11,12,13,14,15 is on GitHub.**

Statistics – Worksheet 1

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

Answer: **A) True**

1. 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**

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

Answer: **B) Modelling bounded count data**

1. Point out the correct statement.

Answer: **D) All of the mentioned**

1. \_\_\_\_\_\_ random variables are used to model rates.

Answer: **C) Poisson**

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

Answer: **B) False**

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

Answer: **B) Hypothesis**

1. Normalized data are centred at\_\_\_\_\_\_and have units equal to standard deviations of the original data.

Answer: **A) 0**

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

Answer: **C) Outliers cannot conform to the regression relationship**

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

* **Normal Distribution is a possibility distribution where the random values are distributed symmetrically. Normal distribution is symmetric from the peak of the curve and it is also known as bell shape curve**. **Mean and median are equal and both located at the center of the distribution. Left side is similar to the right side curve and it appears like mirror image.**

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

**Analyse data with missing values carefully to understand the reasons behind the missing values as it is important to find out how to handle missing values. Another way is to ignore the missing value if the count is less.**

* **Mean imputation**

**Calculate to mean of the non-missing values. It maintain same mean and same sample size.**

* **Substitution**

**When the percentage is large and also avoid bias modelling results,**

**Commonly used ways are mean, median etc. This method could cause bias distribution and variance.**

1. What is A/B testing?

**A/B testing is known as split testing. This is the process of comparing two different versions of data and measure the difference in performance. Give one version to one group and 2nd version to another group check how variation perform. Showing two versions of the asset against one another to make decisions as per the outcome.**

1. Is mean imputation of missing data acceptable practice?

**The process of replacing null values in a data collection with the data’s mean is known as mean imputation. 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.**

1. What is linear regression in statistics?

**Linear regression in statistics is predictive analysis. There are two things to check, First, check if the set of predictor variable does a good job in predicting a result. Second, Check which variables are predictors of the result data. Regression is used to explain the relationship between one dependent variable and one or more independent variable.**

**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.**

1. What are the various branches of statistics?

* **Data Collection.**
* **Descriptive Statistics.**
* **Frequency Distribution.**
* **Discrete and Continuous data.**
* **Inferential Statistics.**

Machine Learning – Worksheet 1

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**

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

Answer: **A) Linear regression is sensitive to outliers**

1. A line falls from left to right if a slope is \_\_\_\_\_\_?

Answer: **A) Positive**

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

Answer: **C) Both of them**

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

Answer**: C) Low bias and high variance**

1. If output involves label then that model is called as :

Answer: **B) Predictive modal**

1. Lasso and Ridge regression techniques belong to \_\_\_\_\_\_\_\_\_?

Answer: **D) Regularization**

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

Answer: **D) SMOTE**

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

Answer: **A) TPR and FPR**

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

Answer: **B) False**

1. Pick the feature extraction from below:

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

1. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

Answers:

**A) We don’t have to choose the learning rate.**

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

**C) We need to iterate.**

1. Explain the term regularization?

Answer: **Regularization means making things regular and acceptable. Regularization is used to reduce errors to avoid overfitting. When huge datasets are passed to the model for training purpose and model start to train over noisy data as well that is called as overfitting. Finding the patterns in underlying dataset and generalize it to predict the target value but the dataset is inflicted with some random noise which causes overfitting or underfitting.**

1. Which particular algorithms are used for regularization?

Answer:

* **Lasso Regression ( L1 )**
* **Least Absolute Shrinkage and Selection Operator regression which adds penalty to the cost function. This is absolute sum of the coefficients. As the value of coefficients increase from 0 this term will penalize the cause model, to decrease the value of coefficients in order to reduce loss.**
* **Ridge Regression ( L2 )**
* **Ridge regression add penalty term which is equal to the square of the coefficient. L2 term is equal to the square of the magnitude of the coefficient. Coefficient is added to control the penalty term.**

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

Answer:

**Error is the margin error within the model, the Sum of deviation within the regression line. This shows the difference between the theoretical value of the model and the actual observed results. Regression line is used as a analysis when determining the correlation between independent variable and dependent variable.**

**Error present in linear regression equation is the error between actual value and the predicted value.**