PYTHON-WORKSHEET 1

1. %
2. 0
3. 24
4. 2
5. 6
6. The finally block will be executed no matter if the try block raises error or not.
7. It is used to raise an exception.
8. In defining a generator.
9. _abc, abc2
10. Yield, raise.

1. 0(n)
2. Logistic Regression
3. Gradient Descent
4. Lasso
5. Stochastic Gradient Descent
6. True
7.
8. Correlation
9. A.) We don't have to choose the learning rate
   B.) It becomes slow when number of features are very large
10. Linear Regression will have low bias and high variance
11. It discovers casual relationship
12. Linear Regression algorithm with training set have millions of features. We can use batch gradient descent, stochastic gradient descent or mini-batch gradient descent. Gradient descent is the process of minimizing a function by following the gradients of the cost function. It involves knowing the form of the cost as well as the derivative so that from a given point you know the gradient and can move in that direction. **Batch gradient descent** is the gradient in which all the training examples are processed for each iteration. If the number of training examples is large, in this case batch gradient descent is computationally very expensive. Hence, if the number of training examples is large, we do not use batch gradient descent. **Stochastic gradient descent** is the gradient which processes one training example for per iteration. Hence, parameters are updated even after one iteration in which only single example has been processed. This is quite faster than batch gradient descent. **Mini Batch gradient descent** is the type which works faster than both batch gradient descent and stochastic gradient descent. In this type it splits the training dataset into small batches that are used to calculate model error and update coefficients. Mini batch gradient descent seeks to find a balance between the robustness of stochastic gradient descent and the efficiency of batch gradient descent. It is the most common implementation of gradient descent used in the field of deep learning.
13. The methods used for normal equations does not require normalization of features, hence the data is not affected by features in the training set, having different scales. For various gradient descent we use feature scaling, which will help us get better and quicker descent coverage. Gradient descent is affected more by the features of different scales, as the model will take more time to reach the maximum scale. Thus the scaling of features is done to avoid the extra time and problem.

STATISTICS-WORKSHEET 1

1. True
2. Central Limit Theorem
3. Modeling bounded count data
4. All of the mentioned
5. Poisson
6. False
7. Hypothesis
8. 0
9. Outliers cannot conform to the regression relationship
10. Normal Distribution curve, also known Gaussian distribution. It is a probability distribution that is symmetric about mean. In graph form it will appear as a **bell curve.** The precise shape varies according to the distribution of data. In normal distribution mean is zero, standard deviation is one. It has zero skews. In normal distribution 68% of observations are within +/- 1 standard deviation, 95% are within +/- 2 standard deviation and 99.7% are within +/- 3 standard deviation. It is based on **Central Limit Theorem**.
11. There are three types of missing data, 1. Missing completely at random, 2.Missing at random and 3.Missing not at random. Missing data can be handled in two ways: 1. **Imputation** 2. **Deletion.** Talking about deletion it can be done in 3 different ways, Deleting rows or listwise deletion, Pairwise deletion and Deleting columns. For imputation it can be done by General problem further divided to **Categorical** and **Continuous,** another is **Time-Series Problem,** further elaborated as Data without trend and without seasonality, data with trend and without seasonality, data without trend and with seasonality.
12. A/B testing also known as split testing, is a basic randomized control experiment, used to compare two or more versions of a variable to find out which performs better in controlled environment. A/B testing eliminates all the guess-work and enables experience optimizers to make data-backed decisions. **A** refers to 'control' or original testing variable, whereas **B** refers to 'variation' or new version of the original testing variable. A/B testing works on following parameters : 1.Make a Hypothesis 2. Create control group or test group 3. Conduct A/B test and collect data. 1. **Make a Hypothesis :** A hypothesis is a assumption or an idea that is proposed so that it can be tested to see if it is true. In other words hypothesis is a tentative insight into the natural world, a concept that is not verified yet but if true would explain certain facts and phenomena. There are two types of hypothesis, **Null Hypothesis H0** and **Alternative hypothesis H1.** The **null hypothesis** is, one that states sample observations result are purely from chance, there is no difference between control and variant group. The **alternative hypothesis** is one that states that sample observations are influenced by some non-random clause, which states there is difference between control and variant group. Once ready with null and alternative hypothesis, next we decide which sample has to participate in test, having two groups- **Control group** and **Test group.** Further we do random selection of samples called **random sampling.** Random sampling is important as it removes or eliminates bias, because we want results of test to be representative rather than the sample itself. While sampling another important aspect is **sample size.** It is required to mention minimum sample size before conducting test so that we can eliminate **under coverage bias.** Finally we conduct the test and collect the data.
13. Mean imputation of missing data is a acceptable practice, but using mean or inserting 0's in missing data can significantly reduce the model's accuracy and bias the results. Advantages of mean imputation, 1. Missing value do not reduce sample size. 2. It is simply to understand and apply. 3. If the response is MCAR, sample mean of instances is not biased. Disadvantages of mean imputation, 1. Mean imputation leads to bias in multivariate estimation such as correlation or regression. 2. Standard errors and variance of imputed variables are biased.
14. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable we want to predict is called independent variable and the variable we are using to predict the other variable is called dependent variable. Linear regression has an equation of the form $Y=a+bX$, where **X** is independent variable and **Y** is dependent variable. The slope of the line is b, and a is the intercept.

15. The two main branches of statistics are **Descriptive statistics** and **Inferential statistics.** Descriptive statistics focuses on collecting, summarizing and presenting a set of data. Inferential statistics analyses sample data to draw conclusions. Descriptive statistics are broken down into: 1. Measures of frequency, 2. Measures of Central tendency, 3. Measures of Dispersion or Variation and 4. Measures of position. 1. **Measure of frequency** include Count, percent, frequency. We can use it when we have to show, how often a response given. 2. **Measure of Central tendency** includes mean, median and mode. Locates distribution of various points, we can use this when we have to show an average or most common response. 3. **Measure of Dispersion or Variation** includes Range, variance and standard deviation. Identifies range of scores by stating intervals. Range is High/Low points and Variance or Standard deviation is difference between observed score and mean, it is helpful to know when data is so spread that it affects the mean. 4. **Measures of position** includes Percentile ranks and Quartile ranks. Describes how scores fall in relation to one another, we can use when we need to compare scores to normalized scores. Similarly Inferential statistics are broken down into: 1. One sample hypothesis test, 2. Confidence interval, 3. Contingency tables and Chi test, 4. T-test or ANOVA, 5.Pearson Correlation, 6.Bi-variate Regression, 7.Multi-variate Regression. 1. **One sample hypothesis test** is a statistical hypothesis test used to determine whether an unknown sample mean is different from a specific value. 2. **Confidence interval** displays the probability that a parameter will fall between a pair of values around the mean. 3. **Contingency tables and chi-test** used to determine whether there is a statistically significant difference between the expected frequencies and observed frequencies in one or more categories. 4. **T-test or ANOVA**, T-test is used to compare the means of two samples, whereas ANOVA is used to compare the means among three or more group. 5. **Pearson Correlation** is correlation coefficient that represents the relationship between two variables that are measured on the same interval or ratio scale. 6. **Bi-variate Regression** is a simple linear regression model which is used to predict one variable from one other variable. 7. **Multi-variate Regression** is used when more than one variable is used to predict variation in another variable.