Python Worksheet:

1. Which of the following operators is used to calculate remainder in a division?
A) # B) & C) % D) $
Answer: C) %

2. In python 2//3 is equal to?
A) 0.666 B) 0 C) 1 D) 0.67
Answer: B) 0

3. In python, 6<<2 is equal to?
A) 36 B) 10 C) 24 D) 45
Answer: c) 24

4. In python, 6&2 will give which of the following as output?
A) 2 B) True C) False D) 0
Answer: A) 2

5. In python, 6|2 will give which of the following as output?
A) 2 B) 4 C) 0 D) 6
Answer: D) 6

6. What does the finally keyword denotes in python?
A) It is used to mark the end of the code
B) It encloses the lines of code which will be executed if any error occurs while executing the lines of code in the try block.
C) the finally block will be executed no matter if the try block raises an error or not.
D) None of the above
Answer C) the finally block will be executed no matter if the try block raises an error or not.

7. What does raise keyword is used for in python?
A) It is used to raise an exception.
B) It is used to define lambda function
C) it's not a keyword in python.
D) None of the above
Answer A) It is used to raise an exception.

8. Which of the following is a common use case of yield keyword in python?
A) in defining an iterator
B) while defining a lambda function
C) in defining a generator
D) in for loop
Answer: C) in defining a generator

9. Which of the following are the valid variable names?
A) _abc B) 1abc C) abc2 D) None of the above
Answer A) _abc C) abc2

10. Which of the following are the keywords in python?
A) yield B) raise C) look-in D) all of the above
Answer A) yield B) raise

Statistics worksheet:

1. Bernoulli random variables take (only) the values 1 and 0.
a) True b) False
Answer a) True

2. 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?
a) Central Limit Theorem b) Central Mean Theorem c) Centroid Limit Theorem
d) All of the mentioned
Answer a) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution? a) Modeling event/time data b) Modeling bounded count data c) Modeling contingency tables d) All of the mentioned
Answer: b) Modeling bounded count data. Poisson distribution is used for modeling unbounded count data

4. Point out the correct statement.
a) The exponent of a normally distributed random variables follows what is called the log- normal distribution
b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
c) The square of a standard normal random variable follows what is called chi-squared distribution
d) All of the mentioned
Answer d) All of the mentioned

5. _____ random variables are used to model rates.
a) Empirical b) Binomial c) Poisson d) All of the mentioned
Answer c) Poisson random variables

6. Usually replacing the standard error by its estimated value does change the CLT.
a) True b) False
Answer b) False

7. Which of the following testing is concerned with making decisions using data?
a) Probability b) Hypothesis c) Causal d) None of the mentioned
Answer b) Hypothesis

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0 b) 5 c) 1 d) 10
Answer a) 0

9. Which of the following statement is incorrect with respect to outliers?
a) Outliers can have varying degrees of influence
b) Outliers can be the result of spurious or real processes
c) Outliers cannot conform to the regression relationship
d) None of the mentioned
Answer c) Outliers can conform to the regression relationship

10. What do you understand by the term Normal Distribution?
Normal Distribution:-
In general normal distribution is where data is distributed normally and ideal normal distribution is where mean = median = mode
In probability theory, a normal (or Gaussian or Gauss or Laplace–Gauss) distribution is a type of continuous probability distribution for a real-valued random variable.
Formula:
The general form of its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

f(x) = probability density function
σ = standard deviation
μ = mean

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

1) Replace with mean/median

If the data is missing we will take the mean/median of the remaining data and replace it with remaining values
Pro:

1. Prevents data loss
2. Easy to implement

Cons:

1. Has a problem of data leakage
2. Works best with smaller data sets as data increases performance will be poorer

2) Delete rows which has missing values

Pro:

1. Creates Robust Model

Cons:

1. Loss of the information
2. If the percentage of missing values is excessive in comparison then i works poorly

3) in categorical column replace them with most frequent category

Pro:

1. Works well with smaller datasets
2. Prevents data loss

Cons:

1. Works only with categorical variable
2. When data increases performance becomes poorer

4) using ML algorithms

Not all ML algorithms support the missing values but algorithms which are robust to the missing values will support. Algorithms like Knn and naive bayes can support missing values where Knn ignores a column and naive bayes use the prediction. But sklearn in python does not support the missing values for KNN and naive bayes

12. What is A/B testing?

A/B testing is a statistical way of comparing two or more versions to determine not only which version performs better but also to understand if a difference between two versions is statistically significant. Businesses run a/b testing to go with data driven approach

Eg:-

1) netflix changes its movie recommendation thumbnail so that to increase the views of a movie
2) When we click on a youtube add out of 2 when we click 1 it will give as a adds in another website also

Businesses don't know whether they understand the customer or not . Users often don't know why they make the choices they make. So it's better to run experiment rather than believing what they take.

13. Is mean imputation of missing data acceptable practice?

If the data is missing we will take the mean/median of the remaining data and replace it with remaining values

Pro:

1. Prevents data loss
2. Easy to implement

Cons:

1. Has a problem of data leakage
2. Works best with smaller data sets as data increases performance will be poorer

14. What is linear regression in statistics?

Answer. We have two types of variables 1. Dependent variable 2. Independent variable

Here dependent variable is continuous variable which we need to determine

In simple linear regression we have only one independent variable so the formulae will be

$Y = M0 + M1x$

But in multi linear regression we have so many independent variables so the formula will change

$Y = M0 + M1X1 + M2X2 + \ldots\ldots + MnXn$

Y = dependent variable

Mi = regression coefficient

Xi = independent variable

Now consider the normal regression expression i.e.,

$Y = 0.9 + 1.2X1 + 1.5X2 + 7X3 + 1X4$

Here X3 coefficient is much bigger than X4 i.e., X3 >> X4 which implies that the X3 independent value is much more important than X4

15. What are the various branches of statistics?
Answer. There are two branches of statistics 1. Descriptive statistics 2. Inferential statistics

1. Descriptive statistics:
   It is a branch of statistics that is concerned with describing and summarizing data; it is called Descriptive statistics. Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.
2. Inferential statistics:
   It as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.

Machine Learning Worksheet:

1. The computational complexity of linear regression is:
A) $O(n\ 2.4\ )$ B) $O(n)$ C) $O(n\ 2\ )$ D) $O(n\ 3\ )$
Answer: B) O(n)

2. Which of the following can be used to fit non-linear data?
A) Lasso Regression B) Logistic Regression C) Polynomial Regression D) Ridge Regression
Answer: B) Logistic Regression

3. Which of the following can be used to optimize the cost function of Linear Regression?
A) Entropy B) Gradient Descent C) Pasting D) None of the above
Answer: B) Gradient Descent

4. Which of the following methods does not have closed form solution for its coefficients?
A) extrapolation B) Ridge C) Lasso D) Elastic Nets

5. Which gradient descent algorithm always gives optimal solution?
A) Stochastic Gradient Descent B) Mini-Batch Gradient Descent C) Batch Gradient Descent D) All of the above
D) All of the above

6. Generalization error measures how well a model performs on training data.
A) True B) False
Answer: A) True

7. The cost function of linear regression can be given as $J(w0, w1\ ) = 1\ 2m \sum (w0 + w1x\ (i) - y\ (i)\ )\ m\ 2\ i{=}1$. The half term at start is due to:
A) scaling cost function by half makes gradient descent converge faster.
B) presence of half makes it easy to do grid search.
C) it does not matter whether half is there or not.
D) None of the above.
Answer: A) scaling cost function by half makes gradient descent converge faster.

8. Which of the following will have symmetric relation between dependent variable and independent variable?
A) Regression B) Correlation C) Both of them D) None of these
B) correlation

9. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
A) We don’t have to choose the learning rate.
B) It becomes slow when number of features are very large.
C) We need to iterate.
D) It does not make use of dependent variable.
Answer: A) , B)

10. Which of the following statement/s are true if we generated data with the help of polynomial features with 5 degrees of freedom which perfectly fits the data?
A) Linear Regression will have high bias and low variance.
B) Linear Regression will have low bias and high variance.
C) Polynomial with degree 5 will have low bias and high variance.
D) Polynomial with degree 5 will have high bias and low variance.
Answer: A) Linear Regression will have high bias and low variance.

11. Which of the following sentence is false regarding regression?
A) It relates inputs to outputs.
B) It is used for prediction.
C) It discovers causal relationship.
D) No inference can be made from regression line.
Answer: C) , D)

12. Which Linear Regression training algorithm can we use if we have a training set with millions of features?
Answer:
In linear regression let us use Gradient Descent. In Gradient Descent if we have millions of features then it will take more time because it will check all million features which is not effective.

Now use the Stochastic gradient descent. In Stochastic Gradient descent the optimization is simpler because it will take features at random but there are downsides when we take a random feature out of million then the result might be biased which is very far from original. It is possible to get outliers or the features are not in sync with others.

Now use mini batch stochastic gradient descent.Here it will take the batch of the features in random but not single one. So here we reduced our computation from a million to a batch and suppressing the downsides of the Stochastic gradient descent. But it doesn't give a very sharp turn because if a mini batch out of million it still has the downsides like the Stochastic gradient descent.

You could use batch gradient descent, stochastic gradient descent, or mini-batch gradient descent. SGD and MBGD would work the best because neither of them need to load the entire dataset into memory in order to take 1 step of gradient descent. Batch would be ok with the caveat that you have enough memory to load all the data.

The normal equations method would not be a good choice because it is computationally inefficient. The main cause of the computational complexity comes from inverse operation on an (n x n) matrix.

13. Which algorithms will not suffer or might suffer, if the features in the training set have very different scales?
Answer:
The normal equations method does not require normalizing the features, so it remains unaffected by features in the training set having very different scales.
Feature scaling is required for the various gradient descent algorithms. Feature scaling will help gradient descent converge quicker.