ASSIGNMENT -39

MACHINE LEARNING

PLEASE NOTE THAT I HAVE NOT COPIED THE QUESTION, ONLY THE CORRECT ANSWER OPTION IS MENTIONED.

Q 1. A

Q 2. A

Q 3. B

Q 4. C

Q 5. C

Q 6. B

Q7. D

Q 8. D

Q 9. A

Q10. B

Q.11. A, B, C

Q.12. A, B

Q.13. Regularization is a method to adjust the machine learning models in order to minimize the adjusted loss and mitigate the occurrence of over fitting or under fitting. Though the use of regularization we can fit our models on a given test set and reduce the errors in it.

Q.14. There are three main algorithms used for regularization of the model to fit the test data, such as the Ridge Regression, Lasso Regression, and Dropout.

Q.15. An error term in a liner regression model implies the margin of error within the statistical model. This means the sum of deviation within the regression line and thus explains the difference between the theoretical value of the model and the actual observed results.

PYTHON – WORKSHEET 1

Q1. C

Q2. B

Q3. C

Q4. A

Q5. D

Q6. C

Q7. A

Q8. A

Q9. D

10. D

STATISTICS WORKSHEET-1

Q1. A

Q2. D

Q3. B

Q4. D

Q5. C

Q6. B

Q7. B

Q8. A

Q9. C

Q10. Normal Distribution is a probability distribution that is symmetric around its mean, that is data points which are closer to the mean have a higher occurrence than data points that are far from the

mean value. Due to this attribute of the data points, the graph of a normal distribution tends to take the shape of a bell-shaped curve.

Q11. We handle missing data through the following imputation techniques:

a) Mean or Median Imputation – This is when data is randomly missing, we can use list-wise or pair-wise deletion of the missing observations.
b) Multivariate Imputation by Chained Equations – Imputes data on a variable-by-variable basis by specifying an imputation model per variable.
c) Random Forest – It is a non-parametric imputation method applicable to various variable types that works with both data missing at random and not missing at random data.

Q12. A/B testing is a user experience research methodology consisting of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics. This is a methodology to compare two versions of a single variable, mainly by testing a respondent's response to variant A vis-a-vis B, and determining which of the two variants is more effective.

Q13. Mean Imputation of missing data can lead to error and is not an advisable practice, it can also lead to underestimation of the standard deviation and distorts relationship between variables by skewing the estimation of correlation towards zero.

Q14. Linear Regression is a statistical approach to predict the relationship between a dependent and an independent variable. These independent and dependent variables are also called explanatory variables. In case we are dealing with one explanatory variable the process is called simple-linear regression while in case of multiple explanatory variables it is called as multiple linear regression.

Q15. There are three branches of statistics, data collection, descriptive statistics, and inferential statistics.