Machine Learning:

In Q1 to Q11, only one option is correct, choose the correct option:

1. Which of the following methods do we use to find the best fit line for data in Linear Regression? - A

2. Which of the following statement is true about outliers in linear regression? - A

3. A line falls from left to right if a slope is? - B

4. Which of the following will have symmetric relation between dependent variable and independent variable? - B

5. Which of the following is the reason for over fitting condition? - C

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

7. Lasso and Ridge regression techniques belong to _____? - D

8. To overcome with imbalance dataset which technique can be used? - D

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph? - C

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

11. Pick the feature extraction from below: B

In Q12, more than one options are correct, choose all the correct options:

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression? A & B

ASSIGNMENT – 39

MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Ans: The word regularize means to make things regular or acceptable. This is exactly why we use it for. Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting.

14. Which particular algorithms are used for regularization?

Ans: There are three main regularization techniques, namely: Ridge Regression (L2 Norm); Lasso (L1 Norm); Dropout

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

Ans: Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was observed. In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.

Statistics:

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

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

3. Which of the following is incorrect with respect to use of Poisson distribution? - B

4. Point out the correct statement. - D

5. _____ random variables are used to model rates. - C

6. 10. Usually replacing the standard error by its estimated value does change the CLT. - B

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

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data. - A

9. Which of the following statement is incorrect with respect to outliers? - C

WORKSHEET

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Ans: Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

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

Ans: Look at the data carefully and find out the missing values. Based on the credibility of that feature we can further choose to drop the feature or not.

Imputation Techniques : Replacing with mean; replacing with previous or next value; Interpolation

12. What is A/B testing?

Ans: Also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

13. Is mean imputation of missing data acceptable practice?

Ans: Mean imputation reduces the variance of the imputed variables. Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval. Mean imputation does not preserve relationships between variables such as correlations.

14. What is linear regression in statistics?

Ans: 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.

15. What are the various branches of statistics?

Ans: There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.

Python:

Q1 to Q8 have only one correct answer. Choose the correct option to answer your question.

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

2. In python 2//3 is equal to? - B

3. In python, 6<<2 is equal to? - C

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

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

6. What does the finally keyword denotes in python? - C

7. What does raise keyword is used for in python? - A

8. Which of the following is a common use case of yield keyword in python? - D

Q9 and Q10 have multiple correct answers. Choose all the correct options to answer your question.

9. Which of the following are the valid variable names? – A, B & C

10. Which of the following are the keywords in python? – A & B

Q11 to Q15 are programming questions. Answer them in Jupyter Notebook.

11. Write a python program to find the factorial of a number.

12. Write a python program to find whether a number is prime or composite.

13. Write a python program to check whether a given string is palindrome or not.

14. Write a Python program to get the third side of right-angled triangle from two given sides.

15. Write a python program to print the frequency of each of the characters present in each string