

MACHINE LEARNING

- 1) A
- 2) A
- 3) B
- 4) B
- 5) C
- 6) B
- 7) A
- 8) D
- 9) A
- 10) C
- 11) C
- 12) A,B

13) Explain the term regularization?

Regularization refers to techniques that are used to calibrate machine-learning models in order to minimize the adjusted loss function and prevent overfitting or under fitting. Using Regularization, we can fit our machine-learning model appropriately on a given test set and hence reduce the errors in it.

14) Which particular algorithms are used for regularization?

Ans: There are two particular algorithms are used for regularization

- a) Ridge Regression
- b) Lasso

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

Ans: The error term of a regression equation represents all of the variation in the dependent variable differ from actual population data. It can also be a variable which represents how a given statistical model differs from reality.

The error term is denoted by ϵ .

PYTHON – WORKSHEET 1

- 1) C
- 2) A
- 3) C
- 4) A
- 5) D
- 6) B
- 7) A
- 8) C
- 9) B, C
- 10) A, B

STATISTICS WORKSHEET-1

- 1) A
- 2) A
- 3) B
- 4) D
- 5) C
- 6) B
- 7) B
- 8) A
- 9) C

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

Ans: In a normal distribution, data are symmetrically distributed with no skew. Most values cluster around a central region, with values tapering off as they go further away from the center. The measures of central tendency (mean, mode, and median) are exactly the same in a normal distribution.

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

Ans: Mean, Median and Mode is one of the most common methods of imputing values when dealing with missing data.

In cases where there are a small number of missing observations, we can calculate the mean or median of the existing observations.

12) What is A/B testing?

Ans: This is a method of determining which of two versions of a product feature, web page, campaign element, or other asset performs better for a certain goal

13) Is mean imputation of missing data acceptable practice?

Ans: 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.

14) What is linear regression in statistics?

Ans: Linear regression is a data analysis technique that predicts the value of unknown data by using another related and known data value. It mathematically models the unknown or dependent variable and the known or independent variable as a linear equation.

15) What are the various branches of statistics?

Ans: The two major areas of statistics:

Descriptive statistics, which describes the properties of sample and population data.

Inferential statistics, which uses those properties to test hypotheses and draw conclusions.