

Machine learning Worksheet

Answers:

- 1) A
- 2) A
- 3) B
- 4) B
- 5) C
- 6) B
- 7) D
- 8) D
- 9) A
- 10) B
- 11) B
- 12) A and B
- 13) 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 underfitting
- 14) : Particular algorithms are used for regularization : 1. Ridge (L2) Regularization :
Also known as Ridge Regression, it modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients. Lasso (L1) Regression It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients.
- 15) The error term of a regression equation represents all of the variation in the dependent variable not explained by the weighted independent variables.

Statistics worksheet

Answers:

- 1) A
- 2) A
- 3) B
- 4) D
- 5) C
- 6) B
- 7) B
- 8) A
- 9) C
- 10) 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. In graphical form, the normal distribution appears as a "bell curve".
- 11) Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your

statistical programme will make the decision for you.

12) A/B testing (also known as bucket testing or split-run testing) is a user experience research methodology. A/B tests consist 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. A/B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants is more effective.

13) The process of replacing null values in a data collection with the data's mean is known as mean imputation.

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. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

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

Python Worksheet Answers:

1) C

2) B

3) C

4) A and C

5) A

6) C

7) A

8) D

9) A, B, and C

10) A and B