

Machine Learning

1.A

2.A

3.B

4.A

5.D

6.B

7.D

8.D

9.C

10.A

11.D

12.A,B,D

13 – Regularization is a different techniques and methods used to address the issue of over fitting by reducing the generalization error without affecting the training error much. Regularization is often used as solution to the overfitting problem in ML.

14 – There are 3 main regularization techniques

1 Ridge Regression and also called as L2 norm

Lasso is also called as L1

Dropout

Ridge and Lasso can be used for any algorithms involving weight parameters, including neural nets. Dropout is primarily used in any kind of neural networks e.g. ANN, DNN, CNN or RNN to moderate the learning.

15- The difference between your model estimate and what is actually observed is small. This difference is the error. It is used to account for the difference between what is observed and what your model estimates.

Python -worksheet 1

1. C

2. B

3. C

4. A

5. D

6. C

7. A

8. A

9. A,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. Normal 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 its appears as bell curve. In normal distribution the mean is zero and the standard deviation is 1. The normal distribution model is important in statistics and key to the Central Limit Theorem (CLT).

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.

The following are common methods for imputation techniques

- Mean imputation - Simply calculate the mean of the observed values for that variable for all individuals who are non-missing
- Substitution.
- Hot deck imputation.
- Cold deck imputation.
- Regression imputation.
- Stochastic regression imputation.
- Interpolation and extrapolation.

12. A/B testing is a user experience research methodology 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.

13. Bad practice in general.

If just estimating means mean imputation preserves the mean of the observed data.
Leads to an underestimate of the standard deviation.

Distorts relationships between variables by “pulling” estimates of the correlation towards zero.

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. The two main branches of statistics are descriptive statistics and inferential 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.

Inferential statistics 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.