

MACHINE LEARNING

ANSWERS

1. D
2. B
3. D
4. A
5. C
6. A , D
7. B , C
8. A , B
9. B , D

SUBJECTIVE ANSWERS

10.

The adjusted R-squared compensates for the addition of variables and only increases if the new predictor improves the model beyond what would be obtained on a probability basis.

11.

Like lasso regression, ridge regression imposes similar constraints on the coefficients by introducing a penalty factor. However, while lasso regression

uses the magnitude of the coefficients, ridge regression takes the square.

Peak regression is also known as L2 smoothing.

12.

The variance inflation factor (VIF) is a measure of the degree of collinearity in a regression analysis. Collinearity occurs when there is a correlation between multiple independent variables in a multiple regression model. This can negatively affect the regression results

13.

To ensure that the gradient descent towards the minima is consistent and that the gradient descent steps are updated at the same rate for all objects, we scale the data before inserting it into the model. Similarly scaled objects help the gradient descent to converge to minima more quickly.

14.

In least squares regression (OLS), three statistics are used to assess the fit of the model: R-square, general F-test, and mean square error (RMSE).

WORKSHEET 6 SQL

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

SUBJECTIVE ANSWERS

11. Denormalization is a database optimization technique where we add redundant data to one or more tables.

12. In computer science, a database cursor is a mechanism for navigating through records in a database.

13. It is generally accepted that there are three different types of searches:

- Browsing search Queries.
- Information search query.

– Transactional searches.

14. Constraints are rules applied to columns of data in a table. They are used to restrict the type of data that can be placed in the array.

15. Auto increment allows you to automatically generate a unique number when you insert a new record into the table. This is often a primary key field that should be automatically created when a new record is inserted.

STATISTICS WORKSHEET- 6

1. A

2. A

3. A

4. C

5. A

6. A

7. A

8. B

9. A

10.

Histograms and boxplots are very similar in that they help visualize and describe numeric data. While histograms are better at determining the underlying distribution of data, boxplots are better for comparing multiple datasets than histograms because they are less detailed and take up less space.

11.

The key is choosing indicators that clearly show where you stand in relation to your goals. Good indicators can be improved.

12.

How do you assess the statistical significance of observations?

13.

Exponential distributions - eg. the amount of time that a car battery lasts or the amount of time until an earthquake occurs.

14.

Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.

15.

Likelihood function is a fundamental concept in statistical inference. It indicates how likely a particular population is to produce an observed sample.

