

Introduction to Data Science

Instructor: Daniel D. Gutierrez

MODULE 9 QUIZ

Question 1

Which one of the following describes supervised machine learning?

- (a) The most common type of machine learning that is typically associated with prediction**
- (b) Provides the ability to discover previously unknown patterns in the data
- (c) Deals with unlabeled data sets having no response variables
- (d) All of the above

Question 2

The following describe aspects of using the linear model in R EXCEPT?

- (a) There is an assumption of a linear relationship between the feature variables and response variable
- (b) The formula argument of the `lm()` algorithm in R is an exact representation of the observed data**
- (c) We can use data visualization to better understand the linear relationship
- (d) The most common approach to fitting a linear model is called *least squares*

Question 3

Which of the following plots contribute to the understanding of linear regression?

- (a) Scatterplot along with a regression line
- (b) Residuals plot
- (c) Predicted vs. residual plot
- (d) All of the above**

Question 4

Which one of the following describes multiple linear regression?

- (a) The linear model data computed by the `lm()` algorithm in R differs considerably from the simple linear regression case
- (b) Uses multiple response variables
- (c) Uses the same least squares approach used in the simple linear regression case**
- (d) None of the above

Question 5

Which of the following describes the need to split a data set into a training set and test set when performing supervised machine learning?

- (a) The separation of the data set into a training portion and a test portion is the way the algorithm learns
- (b) The original data set containing known response variables is split into two pieces
- (c) The training set is used to train the algorithm by calculating the coefficients
- (d) You use the trained model on the test set to predict the response variables that are already known
- (e) All of the above**