Introduction to Data Science

Instructor: Daniel D. Gutierrez

MODULE 6 QUIZ

Question 1

Which of one the following are commonly part of the data munging phase of a data science project?

(a) Converting the raw data format to another format, e.g. CSV to JSON

(b) Loading the data set into a SQL database

(c) Transforming the data set into a form more suitable for machine learning algorithms

(d) All of the above

Question 2

The process of "feature engineering" can be characterized by which one of the following?

(a) Coercing values stored in a variable to a more convenient class

(b) Identifying the subset or raw data or transformed data that you want to use in your machine learning algorithm

(c) The R code necessary to read he data set from its raw form

(d) None of the above

Question 3

The following are types of transformations you might use in the data munging phase of a data science project EXCEPT which one?

(a) Reshaping data sets to obtain a more convenient structure

(b) Merging data sets of similar structure

(c) Transform odd representations of dates and times into appropriate R classes

(d) Computing the mean of all numeric variables

Question 4

The process of imputing missing values in a data set is an important part of data munging. Which one of the following best characterizes imputing?

(a) Remove missing values by discarding incomplete observations

(b) Infer missing values based on data from other observations

(c) Replace missing values for numeric or integer variables with 0

(d) None of the above

Question 5

Which one of the following describes the process of "feature scaling?"

(a) Use the `scale()` function in the base R `stats` package to scale numeric values

(b) Feature scaling allows us to normalize values

(c) Can be used to account for the difference in magnitude due to measurement units

(d) All of the above