

1.11 Introduction to Big Data Techniques

Question 1

Which of the following is *most likely* to cause errors in a machine learning (ML) model that is based on unsupervised learning?

- A. The algorithm overfits the data.
- B. The algorithm's training data is biased.
- C. The algorithm uses unlabeled data to find common patterns.

Question 2

A human investment adviser would *least appropriately* use a robo-adviser to:

- A. create passively managed portfolios.
- B. offer wealth management and planning.
- C. gain a greater market share of cost-conscious investors.

Question 3

An analyst wishes to understand sentiment based on remarks during a recent earnings call. Of the following, the *most appropriate* visualization tool for the analyst to use would be a:

- A. heat map.
- B. word cloud.
- C. frequency polygon.

Question 4

A fund manager is preparing a presentation that shows the fund's number of stock holdings per country and also categorizes the stocks by market cap. Which of the following data visualization tools is *most appropriate* to reflect this data?

- A. Tree-map
- B. Histogram
- C. Scatterplot matrix

Question 5

Natural language processing is *most appropriate* for use when:

- A. analysis is needed for large amounts of sensor-based data.
- B. a company globally monitors employee emails for compliance.
- C. the algorithm must generate predictions that are highly accurate.

Question 6

Machine learning may be *most appropriate* for data analysis if:

- A. training data are applied to targeted outputs.
- B. the input data have not been filtered to eliminate bias.
- C. patterns from raw data inputs were obtained through supervised learning.

Question 7

Use of fintech *most appropriately* enhances a company's risk analysis capabilities when:

- A. machine learning analyzes input data for errors.
- B. the data to be analyzed excludes quantitative data.
- C. the risk analysis focuses mainly on market and trading risk rather than operational risk.

Question 8

A corporate banking analyst is evaluating the concentration risk of a bank's lending portfolio jointly across countries and sectors. The analyst calculates the relative value of the loans (by country and sector) in proportion to the portfolio's value. Which of the following data visualization tools is *most appropriate* to reflect this data?

- A. Heat map
- B. Line chart
- C. Scatterplot

Question 9

Big Data *most likely*:

- A. excludes data generated from traditional sources.
- B. excludes unstructured data and includes only structured data.
- C. is typically composed of data sets generated at high velocities.

Question 10

When considering different sources of Big Data, structured data *most likely* includes:

- A. computer spreadsheets of historic company information.
- B. postings by users of consumer products on social media.
- C. point-of-sale information transmitted by cash registers in stores.

Question 11

With respect to machine learning, a model is *most likely* underfitted when:

- A. the underlying data set is too small to analyze.
- B. the model treats true parameters as if they were noise (ie, insignificant datapoints).
- C. the model treats noise (ie, insignificant datapoints) as if they were true parameters.

Question 12

Which of the following *best* describes fintech within the investment management industry? Fintech is:

- A. the financial analysis of technology stocks.
- B. a form of technical analysis used to determine the price of an asset.
- C. the convergence of finance and technology that provides easier access to investment services.

Question 13

With respect to Big Data, data related to the Internet of Things (IOT) is *most likely* generated through:

- A. sensors.
- B. individuals.
- C. business processes.

Question 14

Which of the following statements is *most* accurate regarding machine learning?

- A. A model is underfitted when it treats noise data as true parameters.
- B. Supervised learning involves using algorithms to learn unlabeled data.
- C. Training data sets help identify historical relationships between variables.

Question 15

A machine learning model treats noise in a training data set as valid parameters and subsequently fails when applied in reality. This outcome is *most likely* a result of:

- A. overfitting.
- B. using structured data.
- C. using unstructured data.