Reading 11

INTRODUCTION TO BIG DATA TECHNIQUES

1. (A) The model treats true parameters as noise.

Explanation

Underfitting describes a machine learning model that is not complex enough to describe the data it is meant to analyze. An underfit model treats true parameters as noise and fails to identify the actual patterns and relationships. A model that is overfit (too complex) will tend to identify spurious relationships in the data. Labeling of input data is related to the use of supervised or unsupervised machine learning techniques.

(Module 11.1, LOS 11.b)

2. (C) Volume and velocity.

Explanation

Big Data may be characterized by its volume (the amount of data available), velocity (the speed at which data are communicated), and variety (degrees of structure in which data exist). "Terabyte" is a measure of volume. "Latency" refers to velocity.

(Module 11.1, LOS 11.b)

3. (A) Text analytics.

Explanation

Text analytics, which relates to the analysis of unstructured data in text or voice forms, can be used to analyze the frequency in which a word or words appear in documents. The number of times the words buy and sell appear can be evaluated using text analytics. Algorithmic trading is computerized securities trading based on preset trading rules. Natural language processing uses computers and artificial intelligence (AI) to interpret human language.

(Module 11.1, LOS 11.c)

4. (C) Fintech companies include those that develop technology for the financial services industry.

Explanation

Fintech refers to technological developments with potential applications in financial services, as well as to the industry that develops these technologies. While firms must process an increasing volume of data, a large portion of that data exists in unstructured forms. Automated investment advice is a potential application of fintech.

(Module 11.1, LOS 11.a)

5. (B) Curation focuses on data quality and accuracy through data cleaning.

Explanation

Curation refers to ensuring the quality and accuracy of data. Capture refers to collecting and transforming data in preparation for analysis. Search refers to the ways data will be queried.

(Module 11.1, LOS 11.c)

6. (C) large amounts of quantitative and qualitative data.

Explanation

The techniques (e.g., simulations and scenario analysis) used to assess and manage risk will require large amounts of quantitative and qualitative data. This is particularly true fora large investment company.

(Module 11.1, LOS 11.c)

7. (A) An analyst adjusts daily stock index data from two countries for their different market holidays.

Explanation

Curation is ensuring the quality of data—for example, by adjusting for bad or missing data. Word clouds are a visualization technique. Moving data from a storage medium to where they are needed is referred to as transfer.

(Module 11.1, LOS 11.b)

8. (B) supervised learning.

Explanation

Supervised learning is a machine learning technique in which a machine is given labelled input and output data and then models the output data based on the input data. In unsupervised learning, a machine is given input data in which to identify patterns and relationships, but no output data to model. Deep learning is a technique to identify patterns of increasing complexity, and may use supervised or unsupervised learning.

(Module 11.1, LOS 11.b)

9. (C) computer systems that emulate human thinking.

Explanation

Artificial intelligence refers to computer systems that emulate the functioning of the human mind. Networks of smart devices and buildings are referred to as the Internet of Things. Data science is the field of study concerned with extracting information from data.

(Module 11.1, LOS 11.b)