

Team CSGO Test Bank

A. BIG DATA

Multiple Choice Questions

1. Which type of analysis provides decision support and makes use of the prediction so that the analysis' result can be used? (Thomas)
 - A. Descriptive analysis
 - B. Diagnostic analysis
 - C. Predictive analysis
 - D. Prescriptive analysis**

2. Which statistical method is used to determine if there exists a relationship between variables? (Thomas)
 - A. A/B testing
 - B. Correlation**
 - C. Regression
 - D. Quantitative

3. Which of the following best describes the challenge of "Variety" in Big Data? (Russell)
 - A. Managing the enormous amount of data being generated
 - B. Handling different types and formats of data such as text, images, and databases**
 - C. Processing data quickly in real time or near real time
 - D. Ensuring the data is accurate and trustworthy

4. What is a major security threat to big data during data ingress (entry)? (Russell)
 - A. Ransomware
 - B. Data leakage
 - C. Malicious data injection**
 - D. Data encryption

5. What does the "Velocity" characteristics of big data refer to? (Richard)
 - A. The speed at which data is generated and processed in real time.**
 - B. The volume of data collected over time.
 - C. The diversity of data types and resources.
 - D. The accuracy and reliability of data.

6. How does big data differ from traditional data in terms of variety? (Richard)
 - A. Big data is only structured, while traditional data is unstructured.

B. Traditional data is structured, while big data includes structured, semi-structured, and unstructured data.

C. Both traditional data and big data are structured.

D. Both traditional data and big data are unstructured.

7. A financial institution needs to build a system that can analyze credit card transactions the moment they occur to identify and block fraudulent purchases. Which combination of technologies is most appropriate for the *real-time processing and analysis* core of this system? (NATHAN)

A. Apache Hadoop for data storage and Apache Spark for batch processing the day's transactions.

B. Apache Kafka to ingest the transaction streams and Apache Flink for real-time processing and fraud detection.

C. Apache Superset for visualizing historical transaction data and Apache Hadoop for storing the results.

D. Apache Spark for scheduled batch analytics and Apache Hive for querying large datasets.

8. According to the document, a key disadvantage of batch processing is "Error propagation," where an error can affect an entire dataset, and errors are only recognized after processing is finished. Which characteristic of stream processing directly addresses this specific disadvantage? (NATHAN)

A. Higher operational costs

B. Error recognition and resolution in real-time

C. Dynamic scalability

D. Demanding hardware requirements

9. What is one key way Big Data helps improve manufacturing operations? (An)

A. By automatically hiring new employees

B. By printing labels faster

C. By predicting machine failures before they happen

D. By increasing shelf life of products

10. Which statement best describes a data lake? (Kevin)

A. A system that stores only structured data

B. A repository for raw data of all types

C. A database for small-scale applications

D. A backup for local devices

True or False

1. Ratio data has equal intervals between values but lacks a true zero point. (Thomas)

Answer: False

2. Big data always contains accurate and trustworthy information without errors or duplicates. (Russell)

Answer: False

3. Ethical use of big data includes safeguarding personal information and preventing unauthorized data collection. (Russell)

Answer: True

4. Apache Spark is considered a suitable tool for a project that requires instantaneous, low-latency insights from a continuous data stream. (Nathan)

Answer: False

5. A major advantage of stream processing is that it provides strong historical context for data analysis because it processes and stores all data before analyzing it. (Nathan)

Answer: False

6. The "Value" characteristics of big data refers to the amount of data collected. (Richard)

Answer: False

7. Big Data can optimize supply chains in retail by analyzing customer purchase patterns. (An)

Answer: True

8. In manufacturing, Big Data has no use in quality control or process monitoring. (An)

Answer: False

9. HDFS can recover data if a DataNode fails. (Kevin)

Answer: True

10. Cloud storage cannot handle multimedia files. (Kevin)

Answer: False

Short Answers

1. In what situation would supervised learning be more appropriate than unsupervised learning? (Thomas)

Answer: Supervised learning is better when the objective is classification and there are labeled data. So the desired output is already known; the algorithm can create a relationship or function between inputs and the correct output. For example, detecting spam emails or frauds.

2. What challenges does the 'Velocity' characteristic of Big Data pose for organizations,? (Russell)

Answer: The "Velocity" challenge refers to the high speed at which data is generated, collected, and needs to be processed. Organizations face difficulties managing this rapid data flow, which can overwhelm traditional data processing systems

3. How do NoSQL databases help with unstructured data in big data? (Richard)

Answer: NoSQL databases like MongoDB store unstructured data (text, images) by using flexible schemas, making it easier to manage diverse data types in big data environments..

4. Give one example of how Big Data is used in the retail industry. (An)

Answer: Retailers use Big Data to recommend products based on customers' purchase history and browsing behavior.

5. Give examples of cloud storage services. (Kevin)

Answer: Google Drive / Dropbox / OneDrive / Amazon S3 / iCloud

Word Banks

1. Big Data Analytics uses several techniques to make sense of large and complex datasets. One major technique is _____ (1) analysis, which deals mainly with measurable data such as sales or profits. It focuses on analyzing _____ (2) to identify clear _____ (3) and relationships that can guide business decisions. In contrast, _____ (4) analysis focuses on non-numerical data such as customer opinions, reviews, or interviews. (Thomas)

Word banks: {quantitative, qualitative, statistical, numbers, patterns}

Answer: quantitative, numbers, patterns, statistical

2. Big Data presents significant _____ challenges that organizations must address to protect user information. One common threat is _____, where unauthorized access leads to data breaches exposing sensitive personal data. Another concern is _____, which involves injecting malicious or false data to corrupt analytics outcomes. To mitigate these risks, organizations implement _____ strategies, including encryption, access controls, and monitoring. Additionally, _____ considerations require respecting user privacy and ensuring data is collected and used transparently and ethically.

Word banks: malicious data injection, security, ethical, cybersecurity, data leakage

Answers: Security, data leakage, malicious data injection, cybersecurity, ethical

3. A data engineering team is designing a new platform. They plan to use 1. _____ to ingest live user clickstreams from their website. The raw data will then be stored long-term in 2. _____ due to its cost-effectiveness for massive datasets. For their nightly sales aggregation reports, which are not time-sensitive, they will use the 3. _____ paradigm. However, to power their live dashboard that tracks user activity as it happens, they need the 4. _____ paradigm. Finally, business analysts will use 5. _____ to create and share interactive visualizations from the processed data. (Nathan)

Word Bank: Batch Processing, Stream Processing, Apache Kafka, Apache Hadoop, Apache Superset

Answer Key : Apache Kafka, Apache Hadoop, Batch Processing, Stream Processing, Apache Superset

4. Big data is a complex 1. _____ that are difficult to analyze using 2. _____ data management data and techniques. The foundations of big data are several 3. _____ and technologies that enable the processing, and extraction of value from large volumes of data. The 5 Vs (Volume, Velocity, 4. _____, Veracity, Value) define the key characteristics of big data. (Richard)

Word Bank: {variety, traditional, concepts, datasets}

Answer: datasets, traditional, concepts, variety

5. Big Data enables _____ maintenance in manufacturing by predicting equipment failure. It also supports _____ control through process data. In retail, it helps in _____ forecasting, customer _____ analysis, and _____ recommendation systems. (An)

Word banks: {quality, product, behavior, predictive, demand}

Correct Answers: predictive, quality, demand, behavior, product

6. _____ manages metadata and directories in Hadoop. _____ stores raw data in any format for analysis or machine learning. The three Vs of big data are _____, _____, and _____. _____ is a process involving Extraction, Transformation, and Loading. _____ databases handle unstructured and semi-structured data efficiently. (Kevin)

Word Bank: (HDFS, Data Lake, ETL, NoSQL, Cloud Storage, NameNode, DataNode, Volume, Velocity, Variety)

Answer: NameNode; Data Lake; Volume, Velocity, Variety; ETL; NoSQL

Multiple Answers

1. Which of the following statements about **Support Vector Machines (SVMs)** are correct? (Thomas)

- A. SVM aims to find the hyperplane that best separates different classes in the feature space.
- B. SVMs can only be used for linear classification problems.
- C. The data points that lie closest to the decision boundary are called support vectors.
- D. SVMs can use kernel functions to handle nonlinear data.
- E. SVMs are primarily designed for clustering unlabeled data.

ANSWER: A, C, D

2. A company is building a data platform and has chosen Apache Kafka as its central data ingestion hub. According to the provided text, which THREE of the following statements accurately describe the role or function that Kafka will play in this architecture? (Nathan)

- A. It will decouple data producers from data consumers, acting as a message broker.
- B. It will perform complex, in-memory batch processing on historical data stored in Hadoop.
- C. It will collect and buffer real-time data streams from sources like application logs and IoT sensors.
- D. It is ideal for implementing an event-driven architecture and handling financial transactions.

Answers: A, C, D

3. Which of the following are examples of sources of big data? (Richard)

- A. Social media posts
- B. Customer names and phone numbers
- C. Website clicks
- D. Sales records in spreadsheets
- E. Transaction data from sensors

Answers: A, C, E

4. Which of the following are benefits of Big Data in manufacturing or retail? (Choose all that apply) An

- A. Predictive maintenance
- B. Personalized marketing
- C. Random price inflation
- D. Inventory optimization
- E. Manual paperwork increase

Answers: A, B, D

5. Which are types of NoSQL databases? (Kevin)

- A. Document-based
- B. Key-value
- C. Graph-based
- D. Table-based

Answers: A, B, C

B. Connectionist AI

Multiple Choice Questions

1. In a Convolutional Neural Network (CNN), what is the main purpose of the convolution layer? (Thomas)

- A. To flatten the data into a one-dimensional vector
- B. To extract spatial features and patterns from the input
- C. To store past information
- D. To reduce the spatial size of feature maps

2. What problem does Backpropagation Through Time (BPTT) aim to solve? (Thomas)

- A. Training RNNs by computing gradients over multiple time steps
- B. Reducing data volumes in CNNs
- C. Increasing the number of convolution kernels
- D. Differentiate whether the image is a dog or a cat

3. What is the primary purpose of an activation function in an artificial neuron? (Nathan)

- A. To calculate the weighted sum of the inputs.
- B. To adjust the learning rate of the network during training.
- C. To introduce non-linearity and map the output to a specific range.
- D. To initialize the weights and biases before training begins.

4. During the aggregation process of a single neuron, if an input has a weight of 0, what is the immediate effect?(Nathan)

- A. The neuron's bias value is automatically set to zero.
- B. The activation function for that neuron fails to compute.
- C. The neuron will ignore that particular input signal.
- D. The output of the neuron is guaranteed to be zero.

5. How does Connectionist AI differ from Symbolic AI in terms of learning? (Richard)

- A. Connectionist AI requires human experts to manually encode rules, while Symbolic AI learns automatically from data.
- B. Connectionist AI learns automatically from examples, while Symbolic AI requires manual rule encoding.
- C. Symbolic AI excels in pattern recognition, while Connectionist AI handles logical reasoning.
- D. Both Connectionist and Symbolic AI store information in human-readable symbols.

6. What is one of the main challenges faced by Connectionist AI models? (Richard)

- A. They can only process structured data.
- B. They require a lot of predefined rules to function.
- C. They are prone to overfitting and generalization issues.
- D. They cannot handle high-dimensional data.

7. Which of the following best describes the black box problem in connectionist AI? (Russell)

A. AI models are entirely interpretable and transparent.

B. The internal workings of neural networks are complex and not easily understood by humans.

C. AI models cannot make predictions from data.

D. Simulation models are less interpretable than AI.

8. What is one future direction of connectionist AI research? (Russell)

A. Reducing the interpretability of models

B. Limiting AI to single-modal tasks

C. Developing neuro-symbolic systems for reasoning and transparency

D. Eliminating on-device computation

9. Which of the following is a real-world example of Connectionist AI in medical imaging analysis? (An)

A. Siri using speech recognition

B. Google Health detecting breast cancer from mammograms

C. Amazon recommending health books

D. Tesla's Autopilot system

10. What is overfitting? (Kevin)

A. When the model performs well on unseen data

B. When the network forgets patterns

C. When the model memorizes training data and fails on new data

D. When data is not trained enough

True or False

1. A convolutional neural network (CNN) is primarily used for analyzing sequential data like text. (Thomas)

Answer: False

2. Without an activation function, a neural network—no matter how many layers—can only model linear relationships. (Nathan)

Answer: True

3. The bias in an artificial neuron acts as a weight for a special, constant input of 1. (Nathan)

Answer: True

4. Connectionist AI learns by adjusting the weights of connections based on the error between predicted and actual outcomes. (Richard)

Answer: True

5. Deep neural networks in connectionist AI are easily interpretable and transparent. (Russell)

Answer: False

6. Connectionist models excel at pattern recognition but often struggle with symbolic reasoning and long-term planning. (Russell)

Answer: True

7. Neural networks can predict the likelihood of developing chronic diseases by analyzing lifestyle and medical data. (An)

Answer: True

8. AI is currently unable to assist in personalizing cancer treatments based on genetic profiles. (An)

Answer: False

9. Generalization refers to how well a model performs on new data. (Kevin)

Answer: True

10. Overfitting improves model accuracy on unseen data. (Kevin)

Answer: False

Short Answers

1. Explain how the architectural differences between Recurrent Neural Networks (RNNs) and Feedforward Neural Networks (FNNs) enable RNNs to perform better on sequential data such as speech or text. (Thomas)

Answer: RNNs have feedback loops that allow information from previous time steps to be used in the current time step in a kind of memory called hidden state. By keeping these memories, it enables them to predict a more accurate outcome based on past information. On the other hand, the FNN process inputs independently without managing the order or context of previous input, making it less effective on sequential data.

2. What is one advantage of connectionist AI over symbolic AI? (Richard)

Answer: Connectionist AI is more adaptable and better at pattern recognition.

3. What is one vulnerability of connectionist AI models? (Russell)

Answer: They are susceptible to adversarial and data poisoning attacks, where small changes in input or training data can lead to incorrect outputs

4. Name one advantage of using AI in drug discovery. (An)

Answer: It speeds up the identification of effective drug compounds and reduces R&D costs.

5. What does "loss function" measure? (Kevin)

Answer: The difference between predicted and actual outputs.

Word Banks

1. A _____ neural network is designed to process data with a grid-like matrix, such as images, using filters/kernels. In this type of neural network, a layer called _____ is used to reduce the data volume and lower the computation workload. (Thomas)

Word Bank: {convolutional, recurrent, feedforward, pooling, activation}

Answer: convolutional, pooling

2. 1. _____ AI stores information in human-readable symbols and logical statements, making it easy to audit and modify. Connectionist systems, however, distribute knowledge across networks of artificial neurons, making them more 2. _____ but harder to interpret. While Symbolic AI excels in tasks that require explicit knowledge and logical reasoning, Connectionist AI shines in 3. _____, adaptive learning, and handling large amounts of data. (Richard)

Word bank: {pattern recognition, symbolic, adaptable}

Answer: symbolic, adaptable, pattern recognition

3. Connectionist AI models require large amounts of _____ to train effectively, which can make them resource-intensive. One of their strengths is the ability to adapt and improve through _____, allowing them to handle complex tasks like _____ and _____. However, these models can be vulnerable to _____ attacks, where small changes in input data can lead to incorrect outputs.

Word Bank: recognition adversarial, data, image recognition speech, learning

Answer: data, learning, image recognition, speech recognition, adversarial (Russell)

4. An artificial neuron, the fundamental unit of a neural network, processes information in a structured way. It begins by receiving multiple input signals. Each input is multiplied by its corresponding strength value, known as (1) _____. The neuron then sums these weighted inputs along with a tunable offset called the (2) _____ in a step known as (3) _____. This resulting sum is then passed to a non-linear (4) _____, such as ReLU or Sigmoid, which decides the final (5) _____ of the neuron, determining what signal is passed on to the next layer in the network. (Nathan)

Word Bank: Weights, Activation Function, Output, Bias, Aggregation

Weights, Bias, Aggregation, Activation Function, Output

5. In the healthcare sector, neural networks help predict _____ risk and detect _____ from scans. In finance, they are used to assess _____ scoring, prevent _____ activity, and support _____ management. (An)

Word bank: {disease, credit, risk, fraudulent, tumors}

Answers: disease, tumors, credit, fraudulent, risk

6. _____ are the connections between nodes that determine signal strength. The process of adjusting weights to reduce errors is called _____. The ability of a network to perform well on unseen data is called _____. When a model learns too much detail from training data, it causes _____. _____ adds a penalty to prevent the model from becoming too complex. (Kevin)

Word Bank: (Input Layer, Hidden Layer, Output Layer, Backpropagation, Generalization, Overfitting, Regularization, Loss Function, Neurons, Weights)

Answer: Weights; Backpropagation; Generalization; Overfitting; Regularization

Multiple Answers

1. Which of the following are examples of connectionist AI architectures? (Thomas)

- A. Convolutional Neural Network (CNN)
- B. Circulating neural network (CNN)
- C. Recurrent Neural Network (RNN)
- D. Feedforward Neural Network (FNN)

Answer: A, C, and D

2. Which of the following are key features of Connectionist AI? (Richard)

- A. It learns from data.
- B. It requires manual rule coding.
- C. It uses activation functions.
- D. It represents knowledge through symbols.

Answer: A, C

3. Which of the following statements are key purposes of an activation function in an artificial neural network? (Nathan)

- A. It determines the final prediction at the output layer of the network.
- B. It introduces non-linearity, allowing the network to learn complex patterns.
- C. It initializes the optimal values for all weights before training.
- D. It maps the neuron's output to a specific, desirable range (e.g., 0 to 1).

Answer: B, D, A

4. Which of the following are valid use cases of Connectionist AI in either Healthcare or Finance? (Select all that apply) (An)

- A. Tumor detection from MRI scans
- B. Credit risk prediction
- C. Ad targeting in social media
- D. Personalized cancer therapy
- E. Forecasting shopping trends

Answers: A, B, D

5. Which are examples of neural network applications? (Kevin)

- A. Speech recognition
- B. Fraud detection
- C. Face recognition
- D. Manual data entry

Answers: A, B, C