

## Exam Questions MLS-C01

AWS Certified Machine Learning - Specialty

<https://www.2passeasy.com/dumps/MLS-C01/>



#### NEW QUESTION 1

A Machine Learning Specialist wants to determine the appropriate SageMakerVariant Invocations Per Instance setting for an endpoint automatic scaling configuration. The Specialist has performed a load test on a single instance and determined that peak requests per second (RPS) without service degradation is about 20 RPS. As this is the first deployment, the Specialist intends to set the invocation safety factor to 0.5. Based on the stated parameters and given that the invocations per instance setting is measured on a per-minute basis, what should the Specialist set as the sageMakervariantinvocationsPerinstance setting?

- A. 10
- B. 30
- C. 600
- D. 2,400

**Answer: C**

#### NEW QUESTION 2

An Machine Learning Specialist discover the following statistics while experimenting on a model.

Experiment 1  
Baseline model  
Train error = 5%  
Test error = 16%

Experiment 2  
The Specialist added more layers and neurons to the model and received the following results:  
Train error = 5.2%  
Test error = 15.7%

Experiment 3  
The Specialist reverted back to the original number of neurons from Experiment 1 and implemented regularization in the neural network, which yielded the following results:  
Train error = 4.7%  
Test error = 9.5%

What can the Specialist learn from the experiments?

- A. The model in Experiment 1 had a high variance error that was reduced in Experiment 3 by regularization. Experiment 2 shows that there is minimal bias error in Experiment 1.
- B. The model in Experiment 1 had a high bias error that was reduced in Experiment 3 by regularization. Experiment 2 shows that there is minimal variance error in Experiment 1.
- C. The model in Experiment 1 had a high bias error and a high variance error that were reduced in Experiment 3 by regularization. Experiment 2 shows that high bias cannot be reduced by increasing layers and neurons in the model.
- D. The model in Experiment 1 had a high random noise error that was reduced in Experiment 3 by regularization. Experiment 2 shows that random noise cannot be reduced by increasing layers and neurons in the model.

**Answer: C**

#### NEW QUESTION 3

A Machine Learning Specialist deployed a model that provides product recommendations on a company's website. Initially, the model was performing very well and resulted in customers buying more products on average. However, within the past few months, the Specialist has noticed that the effect of product recommendations has diminished and customers are starting to return to their original habits of spending less. The Specialist is unsure of what happened, as the model has not changed from its initial deployment over a year ago. Which method should the Specialist try to improve model performance?

- A. The model needs to be completely re-engineered because it is unable to handle product inventory changes.
- B. The model's hyperparameters should be periodically updated to prevent drift.
- C. The model should be periodically retrained from scratch using the original data while adding a regularization term to handle product inventory changes.
- D. The model should be periodically retrained using the original training data plus new data as product inventory changes.

**Answer: D**

#### NEW QUESTION 4

A Machine Learning Specialist was given a dataset consisting of unlabeled data. The Specialist must create a model that can help the team classify the data into different buckets. What model should be used to complete this work?

- A. K-means clustering
- B. Random Cut Forest (RCF)
- C. XGBoost
- D. BlazingText

**Answer: A**

#### NEW QUESTION 5

A Machine Learning Specialist observes several performance problems with the training portion of a machine learning solution on Amazon SageMaker. The solution uses a large training dataset (2 TB in size) and is using the SageMaker k-means algorithm. The observed issues include the unacceptable length of time it takes before the training job launches and poor I/O throughput while training the model. What should the Specialist do to address the performance issues with the current solution?

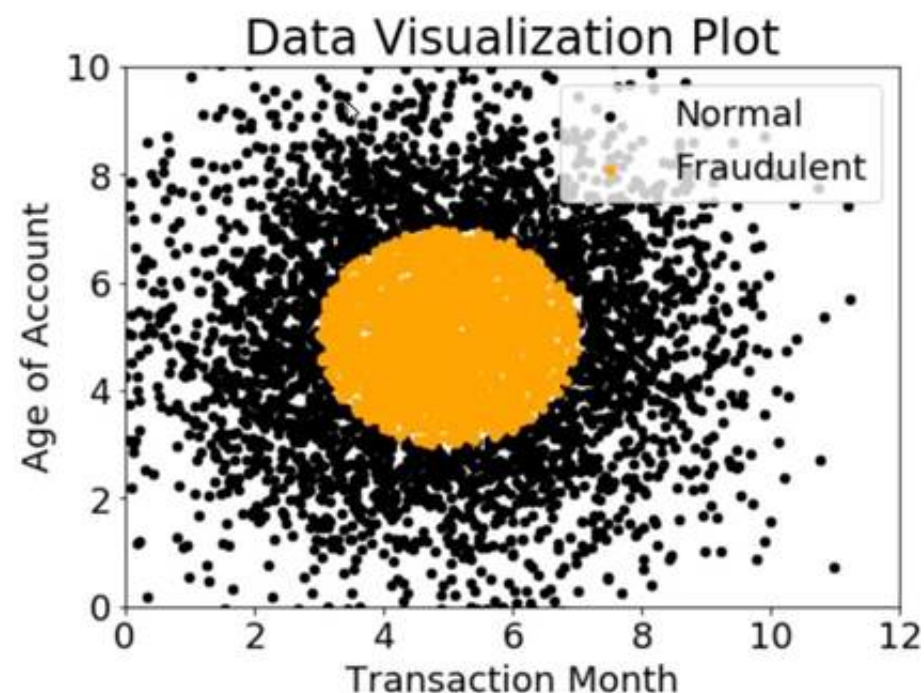
- A. Use the SageMaker batch transform feature.

- B. Compress the training data into Apache Parquet format.
- C. Ensure that the input mode for the training job is set to Pipe.
- D. Copy the training dataset to an Amazon EFS volume mounted on the SageMaker instance.

**Answer: B**

#### NEW QUESTION 6

A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.



Based on this information, which model would have the HIGHEST recall with respect to the fraudulent class?

- A. Decision tree
- B. Linear support vector machine (SVM)
- C. Naive Bayesian classifier
- D. Single Perceptron with sigmoidal activation function

**Answer: C**

#### NEW QUESTION 7

An agency collects census information within a country to determine healthcare and social program needs by province and city. The census form collects responses for approximately 500 questions from each citizen

Which combination of algorithms would provide the appropriate insights? (Select TWO )

- A. The factorization machines (FM) algorithm
- B. The Latent Dirichlet Allocation (LDA) algorithm
- C. The principal component analysis (PCA) algorithm
- D. The k-means algorithm
- E. The Random Cut Forest (RCF) algorithm

**Answer: CD**

#### Explanation:

The PCA and K-means algorithms are useful in collection of data using census form.

#### NEW QUESTION 8

A Machine Learning Specialist is designing a system for improving sales for a company. The objective is to use the large amount of information the company has on users' behavior and product preferences to predict which products users would like based on the users' similarity to other users.

What should the Specialist do to meet this objective?

- A. Build a content-based filtering recommendation engine with Apache Spark ML on Amazon EMR.
- B. Build a collaborative filtering recommendation engine with Apache Spark ML on Amazon EMR.
- C. Build a model-based filtering recommendation engine with Apache Spark ML on Amazon EMR.
- D. Build a combinative filtering recommendation engine with Apache Spark ML on Amazon EMR.

**Answer: B**

#### Explanation:

Many developers want to implement the famous Amazon model that was used to power the "People who bought this also bought these items" feature on Amazon.com. This model is based on a method called Collaborative Filtering. It takes items such as movies, books, and products that were rated highly by a set of users and recommending them to other users who also gave them high ratings. This method works well in domains where explicit ratings or implicit user actions can be gathered and analyzed.

#### NEW QUESTION 9

A Machine Learning Specialist is packaging a custom ResNet model into a Docker container so the company can leverage Amazon SageMaker for training The Specialist is using Amazon EC2 P3 instances to train the model and needs to properly configure the Docker container to leverage the NVIDIA GPUs

What does the Specialist need to do1?

- A. Bundle the NVIDIA drivers with the Docker image

- B. Build the Docker container to be NVIDIA-Docker compatible
- C. Organize the Docker container's file structure to execute on GPU instances.
- D. Set the GPU flag in the Amazon SageMaker Create TrainingJob request body

**Answer:** A

#### NEW QUESTION 10

A Machine Learning Specialist is using Apache Spark for pre-processing training data. As part of the Spark pipeline, the Specialist wants to use Amazon SageMaker for training a model and hosting it. Which of the following would the Specialist do to integrate the Spark application with SageMaker? (Select THREE )

- A. Download the AWS SDK for the Spark environment
- B. Install the SageMaker Spark library in the Spark environment.
- C. Use the appropriate estimator from the SageMaker Spark Library to train a model.
- D. Compress the training data into a ZIP file and upload it to a pre-defined Amazon S3 bucket.
- E. Use the `sageMakerMode`
- F. transform method to get inferences from the model hosted in SageMaker
- G. Convert the DataFrame object to a CSV file, and use the CSV file as input for obtaining inferences from SageMaker.

**Answer:** DEF

#### NEW QUESTION 10

A monitoring service generates 1 TB of scale metrics record data every minute. A Research team performs queries on this data using Amazon Athena. The queries run slowly due to the large volume of data, and the team requires better performance. How should the records be stored in Amazon S3 to improve query performance?

- A. CSV files
- B. Parquet files
- C. Compressed JSON
- D. RecordIO

**Answer:** B

#### NEW QUESTION 12

A Machine Learning Specialist is building a logistic regression model that will predict whether or not a person will order a pizza. The Specialist is trying to build the optimal model with an ideal classification threshold.

What model evaluation technique should the Specialist use to understand how different classification thresholds will impact the model's performance?

- A. Receiver operating characteristic (ROC) curve
- B. Misclassification rate
- C. Root Mean Square Error (RMSE)
- D. L1 norm

**Answer:** A

#### NEW QUESTION 13

A company is running an Amazon SageMaker training job that will access data stored in its Amazon S3 bucket. A compliance policy requires that the data never be transmitted across the internet. How should the company set up the job?

- A. Launch the notebook instances in a public subnet and access the data through the public S3 endpoint
- B. Launch the notebook instances in a private subnet and access the data through a NAT gateway
- C. Launch the notebook instances in a public subnet and access the data through a NAT gateway
- D. Launch the notebook instances in a private subnet and access the data through an S3 VPC endpoint.

**Answer:** D

#### NEW QUESTION 15

A Machine Learning Specialist is building a model to predict future employment rates based on a wide range of economic factors. While exploring the data, the Specialist notices that the magnitude of the input features vary greatly. The Specialist does not want variables with a larger magnitude to dominate the model. What should the Specialist do to prepare the data for model training?

- A. Apply quantile binning to group the data into categorical bins to keep any relationships in the data by replacing the magnitude with distribution
- B. Apply the Cartesian product transformation to create new combinations of fields that are independent of the magnitude
- C. Apply normalization to ensure each field will have a mean of 0 and a variance of 1 to remove any significant magnitude
- D. Apply the orthogonal sparse Diagram (OSD) transformation to apply a fixed-size sliding window to generate new features of a similar magnitude.

**Answer:** C

#### NEW QUESTION 17

A Machine Learning Specialist is working with multiple data sources containing billions of records that need to be joined. What feature engineering and model development approach should the Specialist take with a dataset this large?

- A. Use an Amazon SageMaker notebook for both feature engineering and model development
- B. Use an Amazon SageMaker notebook for feature engineering and Amazon ML for model development
- C. Use Amazon EMR for feature engineering and Amazon SageMaker SDK for model development
- D. Use Amazon ML for both feature engineering and model development.

**Answer:** B



#### NEW QUESTION 19

A Machine Learning Specialist is using Amazon SageMaker to host a model for a highly available customer-facing application . The Specialist has trained a new version of the model, validated it with historical data, and now wants to deploy it to production To limit any risk of a negative customer experience, the Specialist wants to be able to monitor the model and roll it back, if needed What is the SIMPLEST approach with the LEAST risk to deploy the model and roll it back, if needed?

- A. Create a SageMaker endpoint and configuration for the new model versio
- B. Redirect production traffic to the new endpoint by updating the client configuratio
- C. Revert traffic to the last version if the model does not perform as expected.
- D. Create a SageMaker endpoint and configuration for the new model versio
- E. Redirect production traffic to the new endpoint by using a load balancer Revert traffic to the last version if the model does not perform as expected.
- F. Update the existing SageMaker endpoint to use a new configuration that is weighted to send 5% of the traffic to the new varian
- G. Revert traffic to the last version by resetting the weights if the model does not perform as expected.
- H. Update the existing SageMaker endpoint to use a new configuration that is weighted to send 100% of the traffic to the new variant Revert traffic to the last version by resetting the weights if the model does not perform as expected.

**Answer:** A

#### NEW QUESTION 21

A Machine Learning Specialist must build out a process to query a dataset on Amazon S3 using Amazon Athena The dataset contains more than 800.000 records stored as plaintext CSV files Each record contains 200 columns and is approximately 1 5 MB in size Most queries will span 5 to 10 columns only How should the Machine Learning Specialist transform the dataset to minimize query runtime?

- A. Convert the records to Apache Parquet format
- B. Convert the records to JSON format
- C. Convert the records to GZIP CSV format
- D. Convert the records to XML format

**Answer:** A

#### NEW QUESTION 25

An e-commerce company needs a customized training model to classify images of its shirts and pants products The company needs a proof of concept in 2 to 3 days with good accuracy Which compute choice should the Machine Learning Specialist select to train and achieve good accuracy on the model quickly?

- A. . m5 4xlarge (general purpose)
- B. r5.2xlarge (memory optimized)
- C. p3.2xlarge (GPU accelerated computing)
- D. p3 8xlarge (GPU accelerated computing)

**Answer:** C

#### NEW QUESTION 27

A Machine Learning Specialist trained a regression model, but the first iteration needs optimizing. The Specialist needs to understand whether the model is more frequently overestimating or underestimating the target.

What option can the Specialist use to determine whether it is overestimating or underestimating the target value?

- A. Root Mean Square Error (RMSE)
- B. Residual plots
- C. Area under the curve
- D. Confusion matrix

**Answer:** C

#### NEW QUESTION 29

A Machine Learning Specialist is using an Amazon SageMaker notebook instance in a private subnet of a corporate VPC. The ML Specialist has important data stored on the Amazon SageMaker notebook instance's Amazon EBS volume, and needs to take a snapshot of that EBS volume. However the ML Specialist cannot find the Amazon SageMaker notebook instance's EBS volume or Amazon EC2 instance within the VPC.

Why is the ML Specialist not seeing the instance visible in the VPC?

- A. Amazon SageMaker notebook instances are based on the EC2 instances within the customer account, but they run outside of VPCs.
- B. Amazon SageMaker notebook instances are based on the Amazon ECS service within customer accounts.
- C. Amazon SageMaker notebook instances are based on EC2 instances running within AWS service accounts.
- D. Amazon SageMaker notebook instances are based on AWS ECS instances running within AWS service accounts.

**Answer:** C

#### NEW QUESTION 30

A Machine Learning Specialist needs to create a data repository to hold a large amount of time-based training data for a new model. In the source system, new files are added every hour Throughout a single 24-hour period, the volume of hourly updates will change significantly. The Specialist always wants to train on the last 24 hours of the data

Which type of data repository is the MOST cost-effective solution?

- A. An Amazon EBS-backed Amazon EC2 instance with hourly directories
- B. An Amazon RDS database with hourly table partitions
- C. An Amazon S3 data lake with hourly object prefixes
- D. An Amazon EMR cluster with hourly hive partitions on Amazon EBS volumes

Answer: C

#### NEW QUESTION 35

IT leadership wants to transition a company's existing machine learning data storage environment to AWS as a temporary ad hoc solution. The company currently uses a custom software process that heavily leverages SQL as a query language and exclusively stores generated CSV documents for machine learning. The ideal state for the company would be a solution that allows it to continue to use the current workforce of SQL experts. The solution must also support the storage of CSV and JSON files, and be able to query over semi-structured data. The following are high priorities for the company:

- Solution simplicity
- Fast development time
- Low cost
- High flexibility

What technologies meet the company's requirements?

- A. Amazon S3 and Amazon Athena
- B. Amazon Redshift and AWS Glue
- C. Amazon DynamoDB and DynamoDB Accelerator (DAX)
- D. Amazon RDS and Amazon ES

Answer: B

#### NEW QUESTION 37

A Machine Learning Specialist is training a model to identify the make and model of vehicles in images. The Specialist wants to use transfer learning and an existing model trained on images of general objects. The Specialist collated a large custom dataset of pictures containing different vehicle makes and models.

- A. Initialize the model with random weights in all layers including the last fully connected layer.
- B. Initialize the model with pre-trained weights in all layers and replace the last fully connected layer.
- C. Initialize the model with random weights in all layers and replace the last fully connected layer.
- D. Initialize the model with pre-trained weights in all layers including the last fully connected layer.

Answer: B

#### NEW QUESTION 38

For the given confusion matrix, what is the recall and precision of the model?

		Actual	
		Yes	No
Predicted	Yes	12	3
	No	1	9

- A. Recall = 0.92 Precision = 0.84
- B. Recall = 0.84 Precision = 0.8
- C. Recall = 0.92 Precision = 0.8
- D. Recall = 0.8 Precision = 0.92

Answer: A

#### NEW QUESTION 42

A Machine Learning Specialist is assigned a TensorFlow project using Amazon SageMaker for training, and needs to continue working for an extended period with no Wi-Fi access.

Which approach should the Specialist use to continue working?

- A. Install Python 3 and boto3 on their laptop and continue the code development using that environment.
- B. Download the TensorFlow Docker container used in Amazon SageMaker from GitHub to their local environment, and use the Amazon SageMaker Python SDK to test the code.
- C. Download TensorFlow from tensorflow.org to emulate the TensorFlow kernel in the SageMaker environment.
- D. Download the SageMaker notebook to their local environment then install Jupyter Notebooks on their laptop and continue the development in a local notebook.

Answer: A

#### NEW QUESTION 43

.....

## THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual MLS-C01 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the MLS-C01 Product From:

<https://www.2passeasy.com/dumps/MLS-C01/>

## Money Back Guarantee

### MLS-C01 Practice Exam Features:

- \* MLS-C01 Questions and Answers Updated Frequently
- \* MLS-C01 Practice Questions Verified by Expert Senior Certified Staff
- \* MLS-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* MLS-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year