

## Exam Questions MLS-C01

AWS Certified Machine Learning - Specialty

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



#### NEW QUESTION 1

A Machine Learning Specialist is building a convolutional neural network (CNN) that will classify 10 types of animals. The Specialist has built a series of layers in a neural network that will take an input image of an animal, pass it through a series of convolutional and pooling layers, and then finally pass it through a dense and fully connected layer with 10 nodes. The Specialist would like to get an output from the neural network that is a probability distribution of how likely it is that the input image belongs to each of the 10 classes. Which function will produce the desired output?

- A. Dropout
- B. Smooth L1 loss
- C. Softmax
- D. Rectified linear units (ReLU)

**Answer: D**

#### NEW QUESTION 2

A Machine Learning Specialist needs to move and transform data in preparation for training. Some of the data needs to be processed in near-real time and other data can be moved hourly. There are existing Amazon EMR MapReduce jobs to clean and feature engineering to perform on the data. Which of the following services can feed data to the MapReduce jobs? (Select TWO)

- A. AWS DMS
- B. Amazon Kinesis
- C. AWS Data Pipeline
- D. Amazon Athena
- E. Amazon ES

**Answer: BD**

#### NEW QUESTION 3

A Machine Learning team uses Amazon SageMaker to train an Apache MXNet handwritten digit classifier model using a research dataset. The team wants to receive a notification when the model is overfitting. Auditors want to view the Amazon SageMaker log activity report to ensure there are no unauthorized API calls. What should the Machine Learning team do to address the requirements with the least amount of code and fewest steps?

- A. Implement an AWS Lambda function to log Amazon SageMaker API calls to Amazon S3. Add code to push a custom metric to Amazon CloudWatch.
- B. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
- C. Use AWS CloudTrail to log Amazon SageMaker API calls to Amazon S3. Add code to push a custom metric to Amazon CloudWatch.
- D. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
- E. Implement an AWS Lambda function to log Amazon SageMaker API calls to AWS CloudTrail.
- F. Add code to push a custom metric to Amazon CloudWatch.
- G. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
- H. Use AWS CloudTrail to log Amazon SageMaker API calls to Amazon S3. Set up Amazon SNS to receive a notification when the model is overfitting.

**Answer: C**

#### NEW QUESTION 4

A Machine Learning Specialist is preparing data for training on Amazon SageMaker. The Specialist is transformed into a numpy .array, which appears to be negatively affecting the speed of the training. What should the Specialist do to optimize the data for training on SageMaker?

- A. Use the SageMaker batch transform feature to transform the training data into a DataFrame.
- B. Use AWS Glue to compress the data into the Apache Parquet format.
- C. Transform the dataset into the RecordIO protobuf format.
- D. Use the SageMaker hyperparameter optimization feature to automatically optimize the data.

**Answer: C**

#### NEW QUESTION 5

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 6

A Machine Learning Specialist is working for a credit card processing company and receives an unbalanced dataset containing credit card transactions. It contains 99,000 valid transactions and 1,000 fraudulent transactions. The Specialist is asked to score a model that was run against the dataset. The Specialist has been advised that identifying valid transactions is equally as important as identifying fraudulent transactions. What metric is BEST suited to score the model?

- A. Precision
- B. Recall

- C. Area Under the ROC Curve (AUC)
- D. Root Mean Square Error (RMSE)

**Answer:** A

#### NEW QUESTION 7

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 8

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 9

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 10

A Mobile Network Operator is building an analytics platform to analyze and optimize a company's operations using Amazon Athena and Amazon S3. The source systems send data in CSV format in real time. The Data Engineering team wants to transform the data to the Apache Parquet format before storing it on Amazon S3. Which solution takes the LEAST effort to implement?

- A. Ingest .CSV data using Apache Kafka Streams on Amazon EC2 instances and use Kafka Connect S3 to serialize data as Parquet
- B. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Glue to convert data into Parquet.
- C. Ingest .CSV data using Apache Spark Structured Streaming in an Amazon EMR cluster and use Apache Spark to convert data into Parquet.
- D. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Kinesis Data Firehose to convert data into Parquet.

**Answer:** C

#### NEW QUESTION 10

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 12

A company has raw user and transaction data stored in Amazon S3, a MySQL database, and Amazon RedShift. A Data Scientist needs to perform an analysis by joining the three datasets from Amazon S3, MySQL, and Amazon RedShift, and then calculating the average of a few selected columns from the joined data. Which AWS service should the Data Scientist use?

- A. Amazon Athena
- B. Amazon Redshift Spectrum
- C. AWS Glue
- D. Amazon QuickSight

Answer: A

#### NEW QUESTION 16

A Data Scientist is developing a machine learning model to classify whether a financial transaction is fraudulent. The labeled data available for training consists of 100,000 non-fraudulent observations and 1,000 fraudulent observations.

The Data Scientist applies the XGBoost algorithm to the data, resulting in the following confusion matrix when the trained model is applied to a previously unseen validation dataset. The accuracy of the model is 99.1%, but the Data Scientist has been asked to reduce the number of false negatives.

Predicted	0	1
Actual	0 99,966   34	
	1 877   123	

Which combination of steps should the Data Scientist take to reduce the number of false positive predictions by the model? (Select TWO.)

- A. Change the XGBoost eval\_metric parameter to optimize based on rmse instead of error.
- B. Increase the XGBoost scale\_pos\_weight parameter to adjust the balance of positive and negative weights.
- C. Increase the XGBoost max\_depth parameter because the model is currently underfitting the data.
- D. Change the XGBoost eval\_metric parameter to optimize based on AUC instead of error.
- E. Decrease the XGBoost max\_depth parameter because the model is currently overfitting the data.

Answer: DE

#### NEW QUESTION 21

An employee found a video clip with audio on a company's social media feed. The language used in the video is Spanish. English is the employee's first language, and they do not understand Spanish. The employee wants to do a sentiment analysis.

What combination of services is the MOST efficient to accomplish the task?

- A. Amazon Transcribe, Amazon Translate, and Amazon Comprehend
- B. Amazon Transcribe, Amazon Comprehend, and Amazon SageMaker seq2seq
- C. Amazon Transcribe, Amazon Translate, and Amazon SageMaker Neural Topic Model (NTM)
- D. Amazon Transcribe, Amazon Translate, and Amazon SageMaker BlazingText

Answer: C

#### NEW QUESTION 25

A financial services company is building a robust serverless data lake on Amazon S3. The data lake should be flexible and meet the following requirements:

- \* Support querying old and new data on Amazon S3 through Amazon Athena and Amazon Redshift Spectrum.
- \* Support event-driven ETL pipelines.
- \* Provide a quick and easy way to understand metadata. Which approach meets these requirements?

- A. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Glue ETL job, and an AWS Glue Data catalog to search and discover metadata.
- B. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Batch job, and an external Apache Hive metastore to search and discover metadata.
- C. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Batch job, and an AWS Glue Data Catalog to search and discover metadata.
- D. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Glue ETL job, and an external Apache Hive metastore to search and discover metadata.

Answer: B

#### NEW QUESTION 28

A manufacturing company asks its Machine Learning Specialist to develop a model that classifies defective parts into one of eight defect types. The company has provided roughly 100,000 images per defect type for training. During the initial training of the image classification model, the Specialist notices that the validation accuracy is 80%, while the training accuracy is 90%. It is known that human-level performance for this type of image classification is around 90%.

What should the Specialist consider to fix this issue?

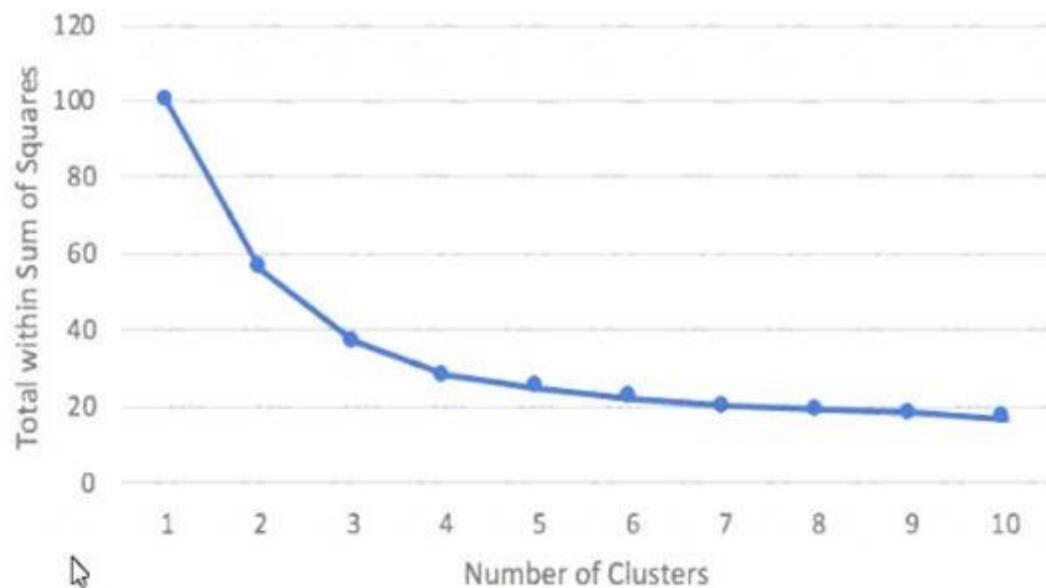
- A. A longer training time
- B. Making the network larger
- C. Using a different optimizer
- D. Using some form of regularization

Answer: D

#### NEW QUESTION 31

A Machine Learning Specialist prepared the following graph displaying the results of k-means for  $k = [1:10]$

### Optimal Number of Clusters



Considering the graph, what is a reasonable selection for the optimal choice of k?

- A. 1
- B. 4
- C. 7
- D. 10

Answer: C

**NEW QUESTION 34**

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 38**

A Machine Learning Specialist is working with a media company to perform classification on popular articles from the company's website. The company is using random forests to classify how popular an article will be before it is published. A sample of the data being used is below. Given the dataset, the Specialist wants to convert the Day-Of-Week column to binary values. What technique should be used to convert this column to binary values?

Article Title	Author	Top Keywords	Day_of_Week	URL_of_Article	Page_Views
Building a Big Data Platform	Jane Doe	Big Data, Spark, Hadoop	Tuesday	<a href="http://examplecorp.com/data_platform.html">http://examplecorp.com/data_platform.html</a>	1300456
Getting Started with Deep Learning	John Doe	Deep Learning, Machine Learning, Spark	Tuesday	<a href="http://examplecorp.com/started_deep_learning.html">http://examplecorp.com/started_deep_learning.html</a>	1230661
MXNet ML Guide	Jane Doe	Machine Learning, MXNet, Logistic Regression	Thursday	<a href="http://examplecorp.com/mxnet_guide.html">http://examplecorp.com/mxnet_guide.html</a>	937291
Intro to NoSQL Databases	Mary Major	NoSQL, Operations, Database	Monday	<a href="http://examplecorp.com/nosql_intro_guide.html">http://examplecorp.com/nosql_intro_guide.html</a>	407812

- A. Binarization
- B. One-hot encoding
- C. Tokenization
- D. Normalization transformation

Answer: B

**NEW QUESTION 42**

A Machine Learning Specialist is working with a large cybersecurity company that manages security events in real time for companies around the world. The cybersecurity company wants to design a solution that will allow it to use machine learning to score malicious events as anomalies on the data as it is being ingested. The company also wants to be able to save the results in its data lake for later processing and analysis. What is the MOST efficient way to accomplish these tasks?

- A. Ingest the data using Amazon Kinesis Data Firehose, and use Amazon Kinesis Data Analytics Random Cut Forest (RCF) for anomaly detection Then use Kinesis Data Firehose to stream the results to Amazon S3
- B. Ingest the data into Apache Spark Streaming using Amazon EMR
- C. and use Spark MLlib with k-means to perform anomaly detection Then store the results in an Apache Hadoop Distributed File System (HDFS) using Amazon EMR with a replication factor of three as the data lake
- D. Ingest the data and store it in Amazon S3 Use AWS Batch along with the AWS Deep Learning AMIs to train a k-means model using TensorFlow on the data in Amazon S3.
- E. Ingest the data and store it in Amazon S3. Have an AWS Glue job that is triggered on demand transform the new data Then use the built-in Random Cut Forest (RCF) model within Amazon SageMaker to detect anomalies in the data

**Answer: B**

#### NEW QUESTION 44

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 45

A manufacturing company has structured and unstructured data stored in an Amazon S3 bucket A Machine Learning Specialist wants to use SQL to run queries on this data. Which solution requires the LEAST effort to be able to query this data?

- A. Use AWS Data Pipeline to transform the data and Amazon RDS to run queries.
- B. Use AWS Glue to catalogue the data and Amazon Athena to run queries
- C. Use AWS Batch to run ETL on the data and Amazon Aurora to run the queries
- D. Use AWS Lambda to transform the data and Amazon Kinesis Data Analytics to run queries

**Answer: D**

#### NEW QUESTION 48

A Data Scientist is developing a machine learning model to predict future patient outcomes based on information collected about each patient and their treatment plans. The model should output a continuous value as its prediction. The data available includes labeled outcomes for a set of 4,000 patients. The study was conducted on a group of individuals over the age of 65 who have a particular disease that is known to worsen with age. Initial models have performed poorly. While reviewing the underlying data, the Data Scientist notices that, out of 4,000 patient observations, there are 450 where the patient age has been input as 0. The other features for these observations appear normal compared to the rest of the sample population. How should the Data Scientist correct this issue?

- A. Drop all records from the dataset where age has been set to 0.
- B. Replace the age field value for records with a value of 0 with the mean or median value from the dataset.
- C. Drop the age feature from the dataset and train the model using the rest of the features.
- D. Use k-means clustering to handle missing features.

**Answer: A**

#### NEW QUESTION 53

A company is using Amazon Polly to translate plaintext documents to speech for automated company announcements However company acronyms are being mispronounced in the current documents How should a Machine Learning Specialist address this issue for future documents'?

- A. Convert current documents to SSML with pronunciation tags
- B. Create an appropriate pronunciation lexicon.
- C. Output speech marks to guide in pronunciation
- D. Use Amazon Lex to preprocess the text files for pronunciation

**Answer: A**

#### NEW QUESTION 57

.....

## 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