Amazon-Web-Services

Exam Questions MLS-C01

AWS Certified Machine Learning - Specialty

NEW QUESTION 1
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. AWSDMS
B. Amazon Kinesis
C. AWS Data Pipeline
D. Amazon Athena
E. Amazon ES

Answer: BD

NEW QUESTION 2
A Machine Learning Specialist is required to build a supervised image-recognition model to identify a cat. The ML Specialist performs some tests and records the following results for a neural network-based image classifier:
Total number of images available = 1,000 Test set images = 100 (constant test set)
The ML Specialist notices that, in over 75% of the misclassified images, the cats were held upside down by their owners.
Which techniques can be used by the ML Specialist to improve this specific test error?

A. Increase the training data by adding variation in rotation for training images.
B. Increase the number of epochs for model training.
C. Increase the number of layers for the neural network.
D. Increase the dropout rate for the second-to-last layer.

Answer: B

NEW QUESTION 3
Example Corp has an annual sale event from October to December. The company has sequential sales data from the past 15 years and wants to use Amazon ML to predict the sales for this year's upcoming event. Which method should Example Corp use to split the data into a training dataset and evaluation dataset?

A. Pre-split the data before uploading to Amazon S3
B. Have Amazon ML split the data randomly.
C. Have Amazon ML split the data sequentially.
D. Perform custom cross-validation on the data

Answer: C

NEW QUESTION 4
A Data Science team is designing a dataset repository where it will store a large amount of training data commonly used in its machine learning models. As Data Scientists may create an arbitrary number of new datasets every day the solution has to scale automatically and be cost-effective. Also, it must be possible to explore the data using SQL.
Which storage scheme is MOST adapted to this scenario?

A. Store datasets as files in Amazon S3.
B. Store datasets as files in an Amazon EBS volume attached to an Amazon EC2 instance.
C. Store datasets as tables in a multi-node Amazon Redshift cluster.
D. Store datasets as global tables in Amazon DynamoDB.

Answer: A

NEW QUESTION 5
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 from the experiments?

A. The model In Experiment 1 had a high variance error lhat 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 thai 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 Expenment 3 by regularization Expenment 2 shows that random noise cannot be reduced by increasing layers and neurons in the model

Answer: C

NEW QUESTION 6
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 long Amazon SageMaker API calls to Amazon S3. Add code to push a custom metric to Amazon CloudWatc
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 CloudWatc
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 CloudTrai
F. Add code to push a custom metric to Amazon CloudWatc
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.

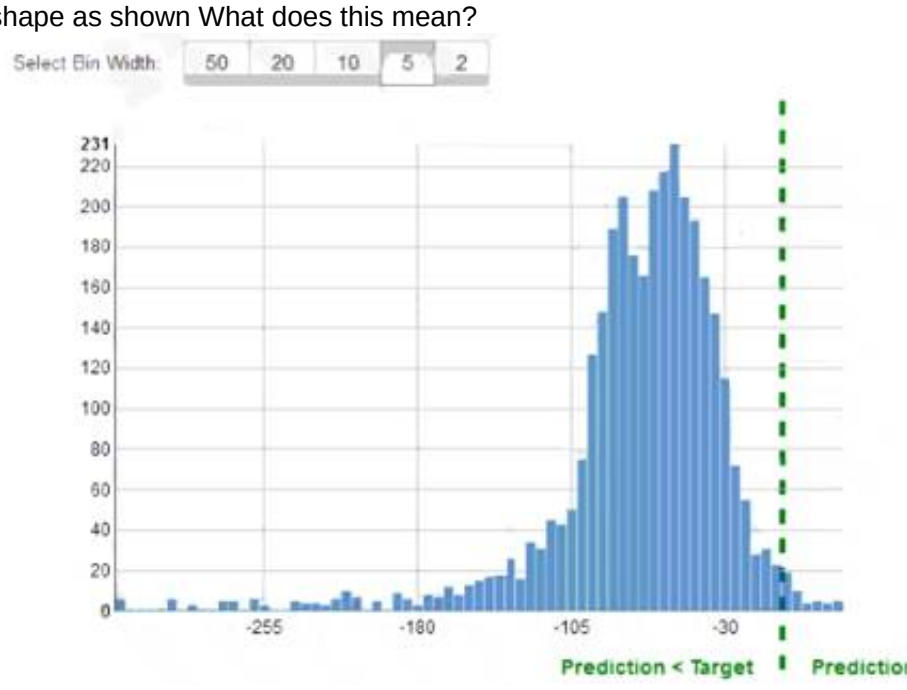
Answer: C

NEW QUESTION 7
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 do?

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 CreateTrainingJob request body

Answer: A

NEW QUESTION 8
While reviewing the histogram for residuals on regression evaluation data a Machine Learning Specialist notices that the residuals do not form a zero-centered bell shape as shown What does this mean?



A. The model might have prediction errors over a range of target values.
B. The dataset cannot be accurately represented using the regression model
C. There are too many variables in the model
D. The model is predicting its target values perfectly.

Answer: D

NEW QUESTION 9
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 10
A large consumer goods manufacturer has the following products on sale
• 34 different toothpaste variants
• 48 different toothbrush variants
• 43 different mouthwash variants
The entire sales history of all these products is available in Amazon S3 Currently, the company is using custom-built autoregressive integrated moving average (ARIMA) models to forecast demand for these products The company wants to predict the demand for a new product that will soon be launched
Which solution should a Machine Learning Specialist apply?

A. Train a custom ARIMA model to forecast demand for the new product.
B. Train an Amazon SageMaker DeepAR algorithm to forecast demand for the new product
C. Train an Amazon SageMaker k-means clustering algorithm to forecast demand for the new product.
D. Train a custom XGBoost model to forecast demand for the new product

Answer: B

Explanation:
The Amazon SageMaker DeepAR forecasting algorithm is a supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNN). Classical forecasting methods, such as autoregressive integrated moving average (ARIMA) or exponential smoothing (ETS), fit a single model to each individual time series. They then use that model to extrapolate the time series into the future.

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 14
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 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 lime 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 16
A retail chain has been ingesting purchasing records from its network of 20,000 stores to Amazon S3 using Amazon Kinesis Data Firehose To support training an improved machine learning model, training records will require new but simple transformations, and some attributes will be combined The model needs lo be retrained daily
Given the large number of stores and the legacy data ingestion, which change will require the LEAST amount of development effort?

A. Require that the stores to switch to capturing their data locally on AWS Storage Gateway for loading into Amazon S3 then use AWS Glue to do the transformation
B. Deploy an Amazon EMR cluster running Apache Spark with the transformation logic, and have the cluster run each day on the accumulating records in Amazon S3, outputting new/transformed records to Amazon S3
C. Spin up a fleet of Amazon EC2 instances with the transformation logic, have them transform the data records accumulating on Amazon S3, and output the transformed records to Amazon S3.
D. Insert an Amazon Kinesis Data Analytics stream downstream of the Kinesis Data Firehouse stream that transforms raw record attributes into simple transformed values using SQL.

Answer: D

NEW QUESTION 20
A Machine Learning Specialist has completed a proof of concept for a company using a small data sample and now the Specialist is ready to implement an end-to-end solution in AWS using Amazon SageMaker The historical training data is stored in Amazon RDS

Which approach should the Specialist use for training a model using that data?

A. Write a direct connection to the SQL database within the notebook and pull data in
B. Push the data from Microsoft SQL Server to Amazon S3 using an AWS Data Pipeline and provide the S3 location within the notebook.
C. Move the data to Amazon DynamoDB and set up a connection to DynamoDB within the notebook to pull data in
D. Move the data to Amazon ElastiCache using AWS DMS and set up a connection within the notebook to pull data in for fast access.

Answer: B

NEW QUESTION 22
A Machine Learning Specialist is building a model that will perform time series forecasting using Amazon SageMaker The Specialist has finished training the model and is now planning to perform load testing on the endpoint so they can configure Auto Scaling for the model variant
Which approach will allow the Specialist to review the latency, memory utilization, and CPU utilization during the load test"?

A. Review SageMaker logs that have been written to Amazon S3 by leveraging Amazon Athena and Amazon OuickSight to visualize logs as they are being produced
B. Generate an Amazon CloudWatch dashboard to create a single view for the latency, memory utilization,and CPU utilization metrics that are outputted by Amazon SageMaker
C. Build custom Amazon CloudWatch Logs and then leverage Amazon ES and Kibana to query and visualize the data as it is generated by Amazon SageMaker
D. Send Amazon CloudWatch Logs that were generated by Amazon SageMaker lo Amazon ES and use Kibana to query and visualize the log data.

Answer: B

NEW QUESTION 26
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 trfese 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 30
A Machine Learning Specialist works for a credit card processing company and needs to predict which transactions may be fraudulent in near-real time. Specifically, the Specialist must train a model that returns the probability that a given transaction may be fraudulent
How should the Specialist frame this business problem'?

A. Streaming classification
B. Binary classification
C. Multi-category classification
D. Regression classification

Answer: A

NEW QUESTION 32
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 34
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	http://examplecorp.com/data_platform.html	1300456
Getting Started with Deep Learning	John Doe	Deep Learning, Machine Learning, Spark	Tuesday	http://examplecorp.com/started_deep_learning.html	1230661
MXNet ML Guide	Jane Doe	Machine Learning, MXNet, Logistic Regression	Thursday	http://examplecorp.com/mxnet_guide.html	937291
Intro to NoSQL Databases	Mary Major	NoSQL, Operations, Database	Monday	http://examplecorp.com/nosql_intro_guide.html	407812

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

Answer: B

NEW QUESTION 35
A Machine Learning Specialist is configuring Amazon SageMaker so multiple Data Scientists can access notebooks, train models, and deploy endpoints. To ensure the best operational performance, the Specialist needs to be able to track how often the Scientists are deploying models, GPU and CPU utilization on the deployed SageMaker endpoints, and all errors that are generated when an endpoint is invoked.
Which services are integrated with Amazon SageMaker to track this information? (Select TWO.)

A. AWS CloudTrail
B. AWS Health
C. AWS Trusted Advisor
D. Amazon CloudWatch
E. AWS Config

Answer: AD

NEW QUESTION 38
A Data Scientist needs to create a serverless ingestion and analytics solution for high-velocity, real-time streaming data.
The ingestion process must buffer and convert incoming records from JSON to a query-optimized, columnar format without data loss. The output datastore must be highly available, and Analysts must be able to run SQL queries against the data and connect to existing business intelligence dashboards.
Which solution should the Data Scientist build to satisfy the requirements?

A. Create a schema in the AWS Glue Data Catalog of the incoming data forma
B. Use an Amazon Kinesis Data Firehose delivery stream to stream the data and transform the data to Apache Parquet or ORC format using the AWS Glue Data Catalog before delivering to Amazon S3. Have the Analysts query the data directly from Amazon S3 using Amazon Athena, and connect to Bl tools using the Athena Java Database Connectivity (JDBC) connector.
C. Write each JSON record to a staging location in Amazon S3. Use the S3 Put event to trigger an AWS Lambda function that transforms the data into Apache Parquet or ORC format and writes the data to a processed data location in Amazon S3. Have the Analysts query the data directly from Amazon S3 using Amazon Athena, and connect to Bl tools using the Athena Java Database Connectivity (JDBC) connector.
D. Write each JSON record to a staging location in Amazon S3. Use the S3 Put event to trigger an AWS Lambda function that transforms the data into Apache Parquet or ORC format and inserts it into an Amazon RDS PostgreSQL databas
E. Have the Analysts query and run dashboards from the RDS database.
F. Use Amazon Kinesis Data Analytics to ingest the streaming data and perform real-time SQL queries to convert the records to Apache Parquet before delivering to Amazon S3. Have the Analysts query the data directly from Amazon S3 using Amazon Athena and connect to Bl tools using the Athena Java Database Connectivity (JDBC) connector.

Answer: A

NEW QUESTION 39
IT leadership wants Jo 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 SOL 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 40
A manufacturer of car engines collects data from cars as they are being driven The data collected includes timestamp, engine temperature, rotations per minute (RPM), and other sensor readings The company wants to predict when an engine is going to have a problem so it can notify drivers in advance to get engine maintenance The engine data is loaded into a data lake for training
Which is the MOST suitable predictive model that can be deployed into production'?

A. Add labels over time to indicate which engine faults occur at what time in the future to turn this into a supervised learning problem Use a recurrent neural network (RNN) to train the model to recognize when an engine might need maintenance for a certain fault.
B. This data requires an unsupervised learning algorithm Use Amazon SageMaker k-means to cluster the data
C. Add labels over time to indicate which engine faults occur at what time in the future to turn this into a supervised learning problem Use a convolutional neural network (CNN) to train the model to recognize when an engine might need maintenance for a certain fault.
D. This data is already formulated as a time series Use Amazon SageMaker seq2seq to model the time series.

Answer: B

NEW QUESTION 44
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 48
A Data Engineer needs to build a model using a dataset containing customer credit card information.
How can the Data Engineer ensure the data remains encrypted and the credit card information is secure? Use a custom encryption algorithm to encrypt the data and store the data on an Amazon SageMaker instance in a VPC. Use the SageMaker DeepAR algorithm to randomize the credit card numbers.

A. Use an IAM policy to encrypt the data on the Amazon S3 bucket and Amazon Kinesis to automatically discard credit card numbers and insert fake credit card numbers.
B. Use an Amazon SageMaker launch configuration to encrypt the data once it is copied to the SageMaker instance in a VP
C. Use the SageMaker principal component analysis (PCA) algorithm to reduce the length of the credit card numbers.
D. Use AWS KMS to encrypt the data on Amazon S3

Answer: C

NEW QUESTION 50
.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

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

100% Actual & Verified — Instant Download, Please Click
Order The MLS-C01 Practice Test Here