

Exam Questions MLS-C01

AWS Certified Machine Learning - Specialty

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



NEW QUESTION 1

A Machine Learning Specialist is implementing a full Bayesian network on a dataset that describes public transit in New York City. One of the random variables is discrete, and represents the number of minutes New Yorkers wait for a bus given that the buses cycle every 10 minutes, with a mean of 3 minutes. Which prior probability distribution should the ML Specialist use for this variable?

- A. Poisson distribution ,
- B. Uniform distribution
- C. Normal distribution
- D. Binomial distribution

Answer: D

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

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 4

A city wants to monitor its air quality to address the consequences of air pollution. A Machine Learning Specialist needs to forecast the air quality in parts per million of contaminants for the next 2 days in the city. As this is a prototype, only daily data from the last year is available.

Which model is MOST likely to provide the best results in Amazon SageMaker?

- A. Use the Amazon SageMaker k-Nearest-Neighbors (kNN) algorithm on the single time series consisting of the full year of data with a predictor_type of regressor.
- B. Use Amazon SageMaker Random Cut Forest (RCF) on the single time series consisting of the full year of data.
- C. Use the Amazon SageMaker Linear Learner algorithm on the single time series consisting of the full year of data with a predictor_type of regressor.
- D. Use the Amazon SageMaker Linear Learner algorithm on the single time series consisting of the full year of data with a predictor_type of classifier.

Answer: C

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 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 6

A gaming company has launched an online game where people can start playing for free but they need to pay if they choose to use certain features. The company needs to build an automated system to predict whether or not a new user will become a paid user within 1 year. The company has gathered a labeled dataset from 1 million users.

The training dataset consists of 1,000 positive samples (from users who ended up paying within 1 year) and 999.1 negative samples (from users who did not use any paid features). Each data sample consists of 200 features including user age, device, location, and play patterns.

Using this dataset for training, the Data Science team trained a random forest model that converged with over 99% accuracy on the training set. However, the prediction results on a test dataset were not satisfactory.

Which of the following approaches should the Data Science team take to mitigate this issue? (Select TWO.)

- A. Add more deep trees to the random forest to enable the model to learn more features.
- B. Indicate a copy of the samples in the test database in the training dataset.
- C. Generate more positive samples by duplicating the positive samples and adding a small amount of noise to the duplicated data.
- D. Change the cost function so that false negatives have a higher impact on the cost value than false positives.
- E. Change the cost function so that false positives have a higher impact on the cost value than false negatives.

Answer: BD

NEW QUESTION 7

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.

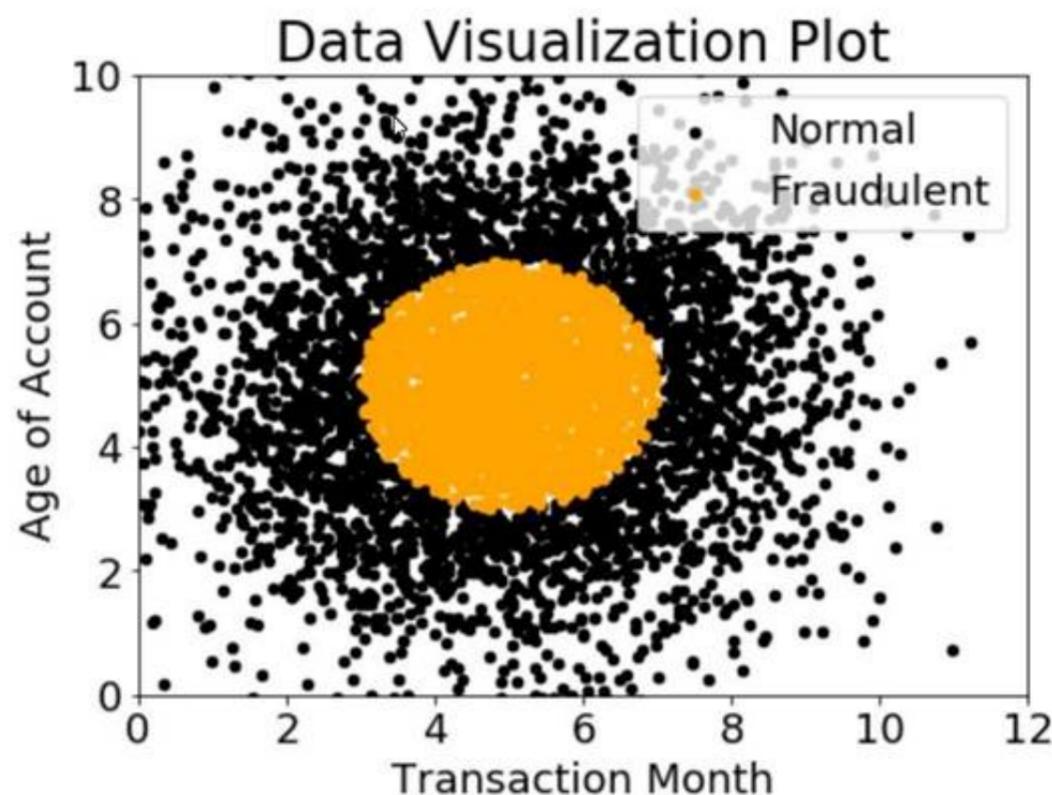
Which method should the Specialist try to improve model performance?

- 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 8

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 accuracy?

- A. Long short-term memory (LSTM) model with scaled exponential linear unit (SELL)
- B. Logistic regression
- C. Support vector machine (SVM) with non-linear kernel
- D. Single perceptron with tanh activation function

Answer: B

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.

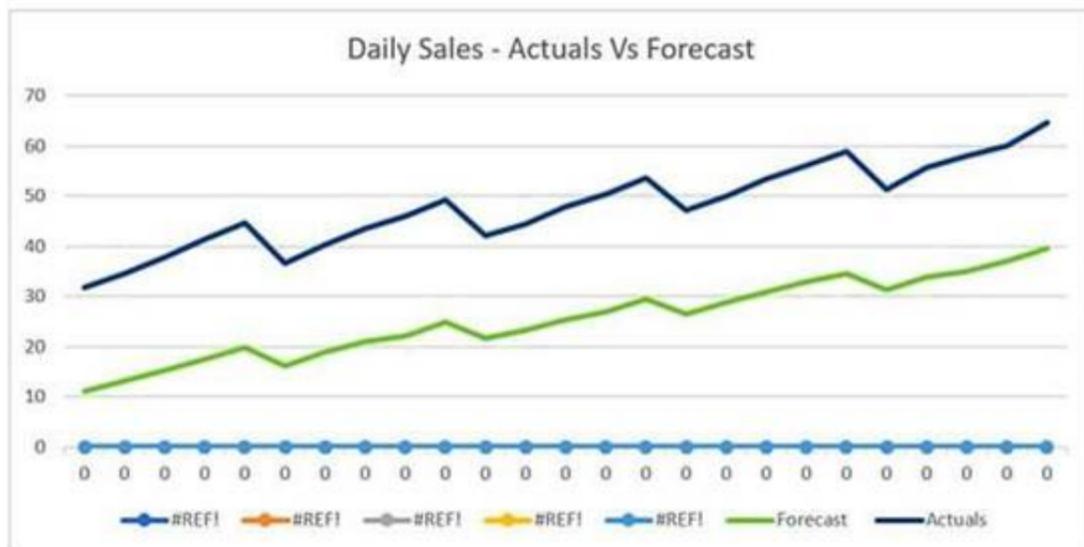
What does the Specialist need to do?

- 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 10

The displayed graph is from a forecasting model for testing a time series.



Considering the graph only, which conclusion should a Machine Learning Specialist make about the behavior of the model?

- A. The model predicts both the trend and the seasonality well.
- B. The model predicts the trend well, but not the seasonality.
- C. The model predicts the seasonality well, but not the trend.
- D. The model does not predict the trend or the seasonality well.

Answer: D

NEW QUESTION 10

A company's Machine Learning Specialist needs to improve the training speed of a time-series forecasting model using TensorFlow. The training is currently implemented on a single-GPU machine and takes approximately 23 hours to complete. The training needs to be run daily. The model accuracy is acceptable, but the company anticipates a continuous increase in the size of the training data and a need to update the model on an hourly, rather than a daily, basis. The company also wants to minimize coding effort and infrastructure changes. What should the Machine Learning Specialist do to the training solution to allow it to scale for future demand?

- A. Do not change the TensorFlow code
- B. Change the machine to one with a more powerful GPU to speed up the training.
- C. Change the TensorFlow code to implement a Horovod distributed framework supported by Amazon SageMaker
- D. Parallelize the training to as many machines as needed to achieve the business goals.
- E. Switch to using a built-in AWS SageMaker DeepAR mode
- F. Parallelize the training to as many machines as needed to achieve the business goals.
- G. Move the training to Amazon EMR and distribute the workload to as many machines as needed to achieve the business goals.

Answer: B

NEW QUESTION 11

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 14

A Machine Learning Specialist is working with a large company to leverage machine learning within its products. The company wants to group its customers into categories based on which customers will and will not churn within the next 6 months. The company has labeled the data available to the Specialist. Which machine learning model type should the Specialist use to accomplish this task?

- A. Linear regression
- B. Classification
- C. Clustering
- D. Reinforcement learning

Answer: B

Explanation:

The goal of classification is to determine to which class or category a data point (customer in our case) belongs to. For classification problems, data scientists

would use historical data with predefined target variables AKA labels (churner/non-churner) – answers that need to be predicted – to train an algorithm. With classification,

businesses can answer the following questions:

- > Will this customer churn or not?
- > Will a customer renew their subscription?
- > Will a user downgrade a pricing plan?
- > Are there any signs of unusual customer behavior?

NEW QUESTION 19

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 21

An insurance company is developing a new device for vehicles that uses a camera to observe drivers' behavior and alert them when they appear distracted The company created approximately 10,000 training images in a controlled environment that a Machine Learning Specialist will use to train and evaluate machine learning models

During the model evaluation the Specialist notices that the training error rate diminishes faster as the number of epochs increases and the model is not accurately inferring on the unseen test images

Which of the following should be used to resolve this issue? (Select TWO)

- A. Add vanishing gradient to the model
- B. Perform data augmentation on the training data
- C. Make the neural network architecture complex.
- D. Use gradient checking in the model
- E. Add L2 regularization to the model

Answer: BD

NEW QUESTION 25

A company is running a machine learning prediction service that generates 100 TB of predictions every day A Machine Learning Specialist must generate a visualization of the daily precision-recall curve from the predictions, and forward a read-only version to the Business team.

Which solution requires the LEAST coding effort?

- A. Run a daily Amazon EMR workflow to generate precision-recall data, and save the results in Amazon S3 Give the Business team read-only access to S3
- B. Generate daily precision-recall data in Amazon QuickSight, and publish the results in a dashboard shared with the Business team
- C. Run a daily Amazon EMR workflow to generate precision-recall data, and save the results in Amazon S3 Visualize the arrays in Amazon QuickSight, and publish them in a dashboard shared with the Business team
- D. Generate daily precision-recall data in Amazon ES, and publish the results in a dashboard shared with the Business team.

Answer: C

NEW QUESTION 29

A company is setting up an Amazon SageMaker environment. The corporate data security policy does not allow communication over the internet.

How can the company enable the Amazon SageMaker service without enabling direct internet access to Amazon SageMaker notebook instances?

- A. Create a NAT gateway within the corporate VPC.
- B. Route Amazon SageMaker traffic through an on-premises network.
- C. Create Amazon SageMaker VPC interface endpoints within the corporate VPC.
- D. Create VPC peering with Amazon VPC hosting Amazon SageMaker.

Answer: A

NEW QUESTION 30

An office security agency conducted a successful pilot using 100 cameras installed at key locations within the main office. Images from the cameras were uploaded to Amazon S3 and tagged using Amazon Rekognition, and the results were stored in Amazon ES. The agency is now looking to expand the pilot into a full production system using thousands of video cameras in its office locations globally. The goal is to identify activities performed by non-employees in real time. Which solution should the agency consider?

- A. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a unique Amazon Kinesis Video Streams video stream
- B. On each stream, use Amazon Rekognition Video and create a stream processor to detect faces from a collection of known employees, and alert when non-

employees are detected.

- C. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a unique Amazon Kinesis Video Streams video stream.
- D. On each stream, use Amazon Rekognition Image to detect faces from a collection of known employees and alert when non-employees are detected.
- E. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camera.
- F. On each stream, use Amazon Rekognition Video and create a stream processor to detect faces from a collection on each stream, and alert when non-employees are detected.
- G. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camera.
- H. On each stream, run an AWS Lambda function to capture image fragments and then call Amazon Rekognition Image to detect faces from a collection of known employees, and alert when non-employees are detected.

Answer: D

NEW QUESTION 34

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` transform method to get inferences from the model hosted in SageMaker.
- F. Convert the DataFrame object to a CSV file, and use the CSV file as input for obtaining inferences from SageMaker.

Answer: DEF

NEW QUESTION 35

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 40

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 45

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 46

A Machine Learning Specialist is developing a custom video recommendation model for an application. The dataset used to train this model is very large with millions of data points and is hosted in an Amazon S3 bucket. The Specialist wants to avoid loading all of this data onto an Amazon SageMaker notebook instance because it would take hours to move and will exceed the attached 5 GB Amazon EBS volume on the notebook instance. Which approach allows the Specialist to use all the data to train the model?

- A. Load a smaller subset of the data into the SageMaker notebook and train locally.
- B. Confirm that the training code is executing and the model parameters seem reasonable.
- C. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
- D. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to the instance.
- E. Train on a small amount of the data to verify the training code and hyperparameter.
- F. Go back to Amazon SageMaker and train using the full dataset.
- G. Use AWS Glue to train a model using a small subset of the data to confirm that the data will be compatible with Amazon SageMaker.
- H. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
- I. Load a smaller subset of the data into the SageMaker notebook and train locally.
- J. Confirm that the training code is executing and the model parameters seem reasonable.
- K. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to train the full dataset.

Answer: A

NEW QUESTION 48

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 49

A Machine Learning Specialist is building a prediction model for a large number of features using linear models, such as linear regression and logistic regression. During exploratory data analysis, the Specialist observes that many features are highly correlated with each other. This may make the model unstable. What should be done to reduce the impact of having such a large number of features?

- A. Perform one-hot encoding on highly correlated features.
- B. Use matrix multiplication on highly correlated features.
- C. Create a new feature space using principal component analysis (PCA).
- D. Apply the Pearson correlation coefficient.

Answer: C

NEW QUESTION 50

A Machine Learning Specialist is building a supervised model that will evaluate customers' satisfaction with their mobile phone service based on recent usage. The model's output should infer whether or not a customer is likely to switch to a competitor in the next 30 days. Which of the following modeling techniques should the Specialist use?

- A. Time-series prediction
- B. Anomaly detection
- C. Binary classification
- D. Regression

Answer: D

NEW QUESTION 52

A Machine Learning Specialist is developing a daily ETL workflow containing multiple ETL jobs. The workflow consists of the following processes:

- * Start the workflow as soon as data is uploaded to Amazon S3.
- * When all the datasets are available in Amazon S3, start an ETL job to join the uploaded datasets with multiple terabyte-sized datasets already stored in Amazon S3.
- * Store the results of joining datasets in Amazon S3.
- * If one of the jobs fails, send a notification to the Administrator. Which configuration will meet these requirements?

- A. Use AWS Lambda to trigger an AWS Step Functions workflow to wait for dataset uploads to complete in Amazon S3. Use AWS Glue to join the datasets. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.
- B. Develop the ETL workflow using AWS Lambda to start an Amazon SageMaker notebook instance. Use a lifecycle configuration script to join the datasets and persist the results in Amazon S3. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.
- C. Develop the ETL workflow using AWS Batch to trigger the start of ETL jobs when data is uploaded to Amazon S3. Use AWS Glue to join the datasets in Amazon S3. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.
- D. Use AWS Lambda to chain other Lambda functions to read and join the datasets in Amazon S3 as soon as the data is uploaded to Amazon S3. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.

Answer: A

NEW QUESTION 56

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 59

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 61

A Data Scientist is working on an application that performs sentiment analysis. The validation accuracy is poor and the Data Scientist thinks that the cause may be a rich vocabulary and a low average frequency of words in the dataset. Which tool should be used to improve the validation accuracy?

- A. Amazon Comprehend syntax analysts and entity detection
- B. Amazon SageMaker BlazingText allow mode
- C. Natural Language Toolkit (NLTK) stemming and stop word removal
- D. Scikit-learn term frequency-inverse document frequency (TF-IDF) vectorizers

Answer: D

NEW QUESTION 66

A Data Scientist wants to gain real-time insights into a data stream of GZIP files. Which solution would allow the use of SQL to query the stream with the LEAST latency?

- A. Amazon Kinesis Data Analytics with an AWS Lambda function to transform the data.
- B. AWS Glue with a custom ETL script to transform the data.
- C. An Amazon Kinesis Client Library to transform the data and save it to an Amazon ES cluster.
- D. Amazon Kinesis Data Firehose to transform the data and put it into an Amazon S3 bucket.

Answer: A

NEW QUESTION 71

An online reseller has a large, multi-column dataset with one column missing 30% of its data. A Machine Learning Specialist believes that certain columns in the dataset could be used to reconstruct the missing data. Which reconstruction approach should the Specialist use to preserve the integrity of the dataset?

- A. Listwise deletion
- B. Last observation carried forward
- C. Multiple imputation
- D. Mean substitution

Answer: C

NEW QUESTION 73

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?

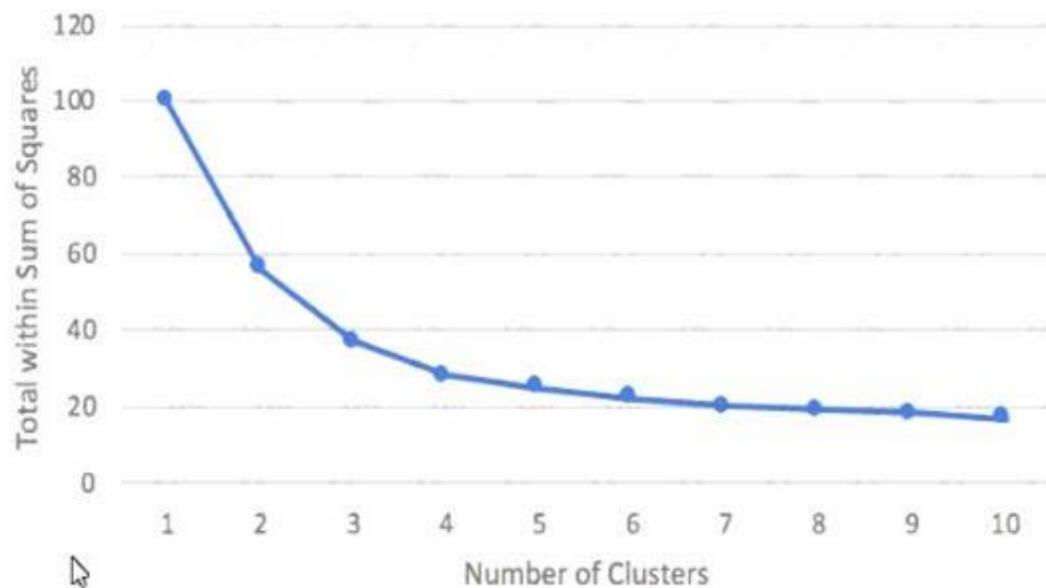
- A. A longer training time
- B. Making the network larger
- C. Using a different optimizer
- D. Using some form of regularization

Answer: D

NEW QUESTION 78

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 83

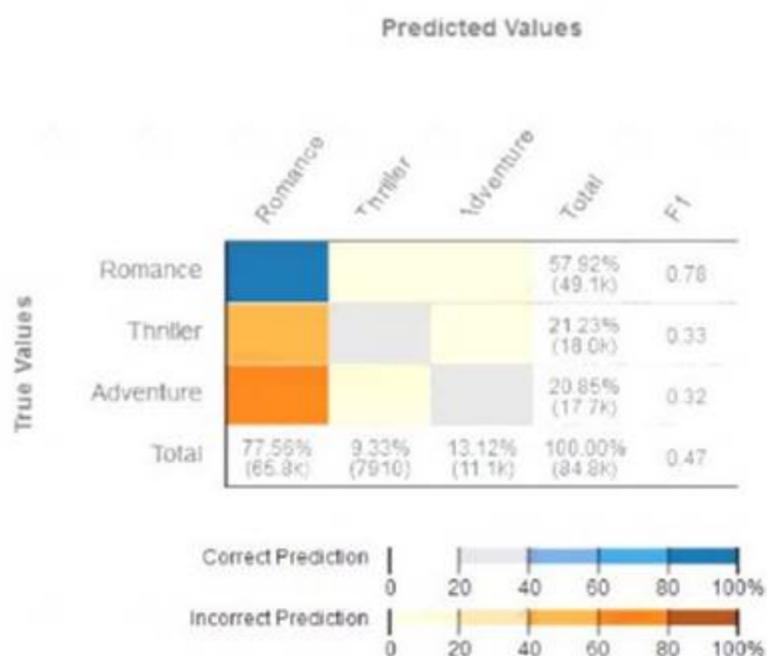
A company is observing low accuracy while training on the default built-in image classification algorithm in Amazon SageMaker. The Data Science team wants to use an Inception neural network architecture instead of a ResNet architecture. Which of the following will accomplish this? (Select TWO.)

- A. Customize the built-in image classification algorithm to use Inception and use this for model training.
- B. Create a support case with the SageMaker team to change the default image classification algorithm to Inception.
- C. Bundle a Docker container with TensorFlow Estimator loaded with an Inception network and use this for model training.
- D. Use custom code in Amazon SageMaker with TensorFlow Estimator to load the model with an Inception network and use this for model training.
- E. Download and apt-get install the inception network code into an Amazon EC2 instance and use this instance as a Jupyter notebook in Amazon SageMaker.

Answer: AD

NEW QUESTION 84

Given the following confusion matrix for a movie classification model, what is the true class frequency for Romance and the predicted class frequency for Adventure?



- A. The true class frequency for Romance is 77.56% and the predicted class frequency for Adventure is 20.85%
- B. The true class frequency for Romance is 57.92% and the predicted class frequency for Adventure is 13.12%
- C. The true class frequency for Romance is 0.78 and the predicted class frequency for Adventure is (0.47 - 0.32).
- D. The true class frequency for Romance is 77.56% * 0.78 and the predicted class frequency for Adventure is 20.85% / 0.32

Answer: A

NEW QUESTION 89

A Machine Learning Specialist needs to be able to ingest streaming data and store it in Apache Parquet files for exploration and analysis. Which of the following services would both ingest and store this data in the correct format?

- A. AWS DMS
- B. Amazon Kinesis Data Streams
- C. Amazon Kinesis Data Firehose
- D. Amazon Kinesis Data Analytics

Answer: C

NEW QUESTION 91

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 96

A Marketing Manager at a pet insurance company plans to launch a targeted marketing campaign on social media to acquire new customers. Currently, the company has the following data in Amazon Aurora:

- Profiles for all past and existing customers
- Profiles for all past and existing insured pets
- Policy-level information
- Premiums received
- Claims paid

What steps should be taken to implement a machine learning model to identify potential new customers on social media?

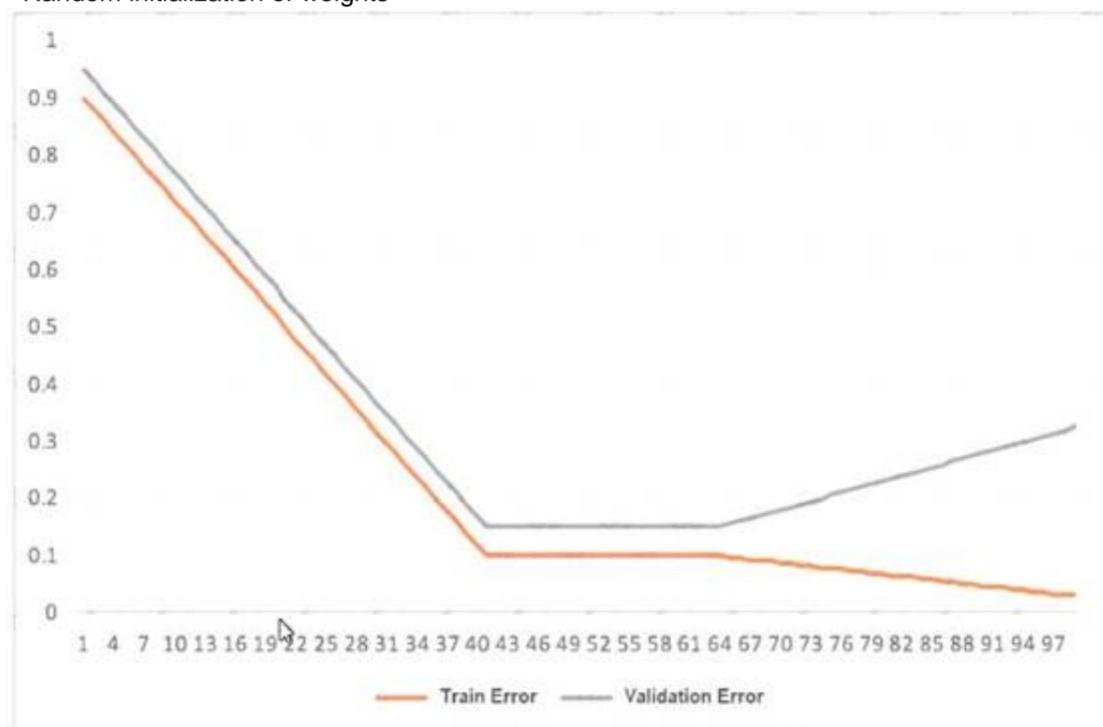
- A. Use regression on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media.
- B. Use clustering on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media.
- C. Use a recommendation engine on customer profile data to understand key characteristics of consumer segments.
- D. Find similar profiles on social media.
- E. Use a decision tree classifier engine on customer profile data to understand key characteristics of consumer segments.
- F. Find similar profiles on social media.

Answer: C

NEW QUESTION 99

This graph shows the training and validation loss against the epochs for a neural network. The network being trained is as follows:

- Two dense layers, one output neuron
- 100 neurons in each layer
- 100 epochs
- Random initialization of weights



Which technique can be used to improve model performance in terms of accuracy in the validation set?

- A. Early stopping
- B. Random initialization of weights with appropriate seed
- C. Increasing the number of epochs
- D. Adding another layer with the 100 neurons

Answer: D

NEW QUESTION 101

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 103

During mini-batch training of a neural network for a classification problem, a Data Scientist notices that training accuracy oscillates. What is the MOST likely cause of this issue?

- A. The class distribution in the dataset is imbalanced
- B. Dataset shuffling is disabled
- C. The batch size is too big
- D. The learning rate is very high

Answer: D

NEW QUESTION 106

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 108

A manufacturing company has a large set of labeled historical sales data. The manufacturer would like to predict how many units of a particular part should be produced each quarter. Which machine learning approach should be used to solve this problem?

- A. Logistic regression
- B. Random Cut Forest (RCF)
- C. Principal component analysis (PCA)
- D. Linear regression

Answer: B

NEW QUESTION 111

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 114

A Machine Learning Specialist has built a model using Amazon SageMaker built-in algorithms and is not getting expected accurate results. The Specialist wants to use hyperparameter optimization to increase the model's accuracy. Which method is the MOST repeatable and requires the LEAST amount of effort to achieve this?

- A. Launch multiple training jobs in parallel with different hyperparameters.
- B. Create an AWS Step Functions workflow that monitors the accuracy in Amazon CloudWatch Logs and relaunches the training job with a defined list of hyperparameters.
- C. Create a hyperparameter tuning job and set the accuracy as an objective metric.
- D. Create a random walk in the parameter space to iterate through a range of values that should be used for each individual hyperparameter.

Answer: B

NEW QUESTION 117

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 122

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 126

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 128

.....

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