

Question #20 Topic 1

An interactive online dictionary wants to add a widget that displays words used in similar contexts. A Machine Learning Specialist is asked to provide word features for the downstream nearest neighbor model powering the widget.

What should the Specialist do to meet these requirements?

A. Create one-hot word encoding vectors.
 B. Produce a set of synonyms for every word using Amazon Mechanical Turk.
 C. Create word embedding vectors that store edit distance with every other word.
 D. Download word embeddings pre-trained on a large corpus.

WORD embeddings

Reveal Solution **Discussion 22**

Question #21 Topic 1

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? (Choose two.)

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

Alls

Logs
CloudWatch Metrics

Reveal Solution **Discussion 0**

Question #22 Topic 1

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 to 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 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 Firehose stream that transforms raw record attributes into simple transformed values using SQL.

Question #23 Topic 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)

Measure performance

Reveal Solution **Discussion 19**

Question #24 Topic 1

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

Measure performance

Reveal Solution **Discussion 20**

Question #25 Topic 1

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.

Data Visualization Plot

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

Search 18:05 Tue 31 Jan 2023 76% **Topic 1**

Home > Amazon Exams > AWS-Certif...

Disclaimer: ExamTopics website is not related to, affiliated with, endorsed by or authorized by Amazon. Examinations, certification & product names are used for reference only and belong to Amazon.

Topic 1 - Single Topic

Topic #1 Topic 1

A large mobile network operating company is building a machine learning model to predict customers who are likely to unsubscribe from the service. The company plans to offer an incentive for these customers as the cost of churn is far greater than the cost of the incentive. The model produces the following confusion matrix after evaluating on a test dataset of 100 customers:

	PREDICTED CHURN	PREDICTED NO CHURN
ACTUAL Churn Yes	10	76
ACTUAL No	10	76

Based on the model evaluation results, why is this a viable model for production?

A. The model is 86% accurate and the cost incurred by the company as a result of false negatives is less than the false positives.
 B. The precision of the model is 86%, which is less than the accuracy of the model.
 C. The model is 86% accurate and the cost incurred by the company as a result of false positives is less than the false negatives.
 D. The precision of the model is 86%, which is greater than the accuracy of the model.

option c

Reveal Solution **Discussion 38**

Question #2 Topic 1

Question #2 Topic 1

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

Question #6 Topic 1

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?

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

Reveal Solution **Discussion 18**

Question #7 Topic 1

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?

- Review SageMaker logs that have been written to Amazon S3 by leveraging Amazon Athena and Amazon QuickSight to visualize logs as they are being produced.
- 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.
- Build custom Amazon CloudWatch Logs and then leverage Amazon ES and Kibana to query and visualize the log data as it is generated by Amazon SageMaker.
- Send Amazon CloudWatch Logs that were generated by Amazon SageMaker to Amazon ES and use Kibana to query and visualize the log data.

Reveal Solution **Discussion 15**

Question #8 Topic 1

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?

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

Reveal Solution **Discussion 16**

Question #9 Topic 1

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?

- Load a smaller subset of the data into the SageMaker notebook and train locally. Confirm that the training code is executing and the model parameters seem reasonable. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
- Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to the instance. Train on a small amount of the data to verify the training code and hyperparameters. Go back to Amazon SageMaker and train using the full dataset.
- 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. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
- Load a smaller subset of the data into the SageMaker notebook and train locally. Confirm that the training code is executing and the model parameters seem reasonable. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to train the full dataset.

Reveal Solution **Discussion 15**

Question #10 Topic 1

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?

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

DATA pipeline for migration inside AWS

Question #26 Topic 1

A Machine Learning Specialist kicks off a hyperparameter tuning job for a tree-based ensemble model using Amazon SageMaker with Area Under the ROC Curve (AUC) as the objective metric. This workflow will eventually be deployed in a pipeline that re-trains and tunes hyperparameters each night to model click-through on data that goes stale every 24 hours. With the goal of decreasing the amount of time it takes to train these models, and ultimately to decrease costs, the Specialist wants to reconfigure the input hyperparameter range(s). Which visualization will accomplish this?

- A histogram showing whether the most important input feature is Gaussian.
- A scatter plot with points colored by target variable that uses t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize the large number of input variables in an easier-to-read dimension.
- A scatter plot showing the performance of the objective metric over each training iteration.
- A scatter plot showing the correlation between maximum tree depth and the objective metric.

Reveal Solution **Discussion 33**

Question #27 Topic 1

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?

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

Reveal Solution **Discussion 51**

Question #4 Topic 1

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?

- 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.
- Use Amazon SageMaker Random Cut Forest (RCF) on the single time series consisting of the full year of data with a predictor_type of regressor.
- 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.
- 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.

**KM → clustering
CWT RF → fraud
linear learner ✓
predictor → regression**

Question #27 Topic 1

A Machine Learning Specialist is creating a new natural language processing application that processes a dataset comprised of 1 million sentences. The aim is to then run Word2Vec to generate embeddings of the sentences and enable different types of predictions. Here is an example from the dataset:
"The quck BROWN FOX jumps over the lazy dog."

Which of the following are the operations the Specialist needs to perform to correctly sanitize and prepare the data in a repeatable manner? (Choose three.)

- Perform part-of-speech tagging and keep the action verb and the nouns only.
- Normalize all words by making the sentence lowercase.
- Remove stop words using an English stopword dictionary.
- Correct the typography on "quck" to "quick."
- One-hot encode all words in the sentence.
- Tokenize the sentence into words.

Reveal Solution **Discussion 17**

Question #28**Topic 1**

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

Question #30**Topic 1**

When submitting Amazon SageMaker training jobs using one of the built-in algorithms, which common parameters MUST be specified? (Choose three.)

- A. The training channel identifying the location of training data on an Amazon S3 bucket.
- B. The validation channel identifying the location of validation data on an Amazon S3 bucket.
- C. The IAM role that Amazon SageMaker can assume to perform tasks on behalf of the users.
- D. Hyperparameters in a JSON array as documented for the algorithm used.
- E. The Amazon EC2 instance class specifying whether training will be run using CPU or GPU.

Question #32**Topic 1**

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.

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	1200456
Getting Started with Deep Learning	John Doe	Deep Learning, Machine Learning, Spark	Tuesday	http://examplecorp.com/started_deep_learning.html	1230061
MXNet ML Guide	Jane Doe	Machine Learning, TensorFlow, 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

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?

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

Reveal Solution**Discussion 10****Question #34****Topic 1**

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

Reveal Solution**Discussion 46****Question #29****Topic 1**

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? (Choose 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.

Question #31**Topic 1**

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

Reveal Solution**Discussion 5****Question #33****Topic 1**

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 999,000 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? (Choose two.)

- A. Add more deep trees to the random forest to enable the model to learn more features.
- B. Include a copy of the samples in the test dataset 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.

Reveal Solution**Discussion 23**

Question #35

Topic 1

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.

Question #37

Topic 1

A Machine Learning Specialist working for an online fashion company wants to build a data ingestion solution for the company's Amazon S3-based data lake.

The Specialist wants to create a set of ingestion mechanisms that will enable future capabilities comprised of:

- ⇒ Real-time analytics
- ⇒ Interactive analytics of historical data
- ⇒ Clickstream analytics
- ⇒ Product recommendations

Which services should the Specialist use?

- A. AWS Glue as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for real-time data insights; Amazon Kinesis Data Firehose for delivery to Amazon ES for clickstream analytics; Amazon EMR to generate personalized product recommendations
- B. Amazon Athena as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for near-real-time data insights; Amazon Kinesis Data Firehose for clickstream analytics; AWS Glue to generate personalized product recommendations
- C. AWS Glue as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for historical data insights; Amazon Kinesis Data Firehose for delivery to Amazon ES for clickstream analytics; Amazon EMR to generate personalized product recommendations
- D. Amazon Athena as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for historical data insights; Amazon DynamoDB streams for clickstream analytics; AWS Glue to generate personalized product recommendations

Catalog - b/c

Question #36

Topic 1

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.

Question #38

Topic 1

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? (Choose 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.

A Machine Learning Specialist built an image classification deep learning model. However, the Specialist ran into an overfitting problem in which the training and testing accuracies were 99% and 75%, respectively.

How should the Specialist address this issue and what is the reason behind it?

- A. The learning rate should be increased because the optimization process was trapped at a local minimum.
- B. The dropout rate at the flatten layer should be increased because the model is not generalized enough.
- C. The dimensionality of dense layer next to the flatten layer should be increased because the model is not complex enough.
- D. The epoch number should be increased because the optimization process was terminated before it reached the global minimum

Question #40

Topic 1

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

QUESTION 1

Topic: Modeling

A Machine Learning Specialist has thousands of customer feedback that are labeled as neutral, positive, and negative. Some feedback contains informal language and slang words. The Specialist needs to come up with a baseline ML-model that analyzes sentiments.

Which machine learning algorithm should the Specialist use?

K-means Clustering X

Multinomial Logistic Regression

Recurrent Neural Network (RNN)

Latent Dirichlet Allocation (LDA) X

QUESTION 2

Topic: Data Engineering

A Machine Learning Specialist has an Amazon S3-based data lake that contains a gigabyte-size worth of training data and their associated metadata. The Specialist needs to perform ad-hoc queries on the metadata to inspect the dataset. The Specialist wants a solution that has the least amount of effort.

How can the Specialist achieve his goal?

Use Amazon S3 Analytics to explore the metadata.

Trigger a Lambda function to download the files into the /tmp directory. Then, scan and search through the metadata.

Search through the metadata using Amazon Athena.

Launch an Amazon EMR with Apache Spark to search through the metadata.

QUESTION 3

Topic: Machine Learning Implementation and Operations

A Machine Learning Specialist wants to test multiple variants of a model hosted in a development environment where it runs inference for real-time data. The Specialist needs to perform A/B testing to identify which model is the best candidate to deploy in production.

Which approach will enable the Specialist to do this with the LEAST amount of configuration effort?

Deploy the model variants behind multiple Amazon SageMaker endpoints. Use an Application Load Balancer to distribute a percentage of traffic to each model. Redirect all traffic to the best performing model once the A/B test is done.

Use API Gateway to an API endpoint for each model variant. Advertise the endpoints for A/B testing. Delete all model endpoints that performed poorly and only retain the API endpoint of the best performing model.

Launch multiple EC2 instances with the AWS Deep Learning AMI (DLAMI) to host different model variants. Run the instances behind an Application Load Balancer.

Deploy the model variants behind a single Amazon SageMaker endpoint. Specify a traffic percentage for each variant and shift all traffic to the best performing model once the A/B test is done.

QUESTION 4

Topic: Machine Learning Implementation and Operations

A Systems Administrator is setting up an Amazon SageMaker environment for a Machine Learning team. The administrator must give each member access to a single Amazon SageMaker notebook instance that they will use exclusively during their term.

How should the Administrator control the user access for each individual SageMaker notebook instance?

Modify the bucket policy of the S3 bucket that holds the training data to only allow members to access their personal SageMaker notebook instance.

Create an IAM policy that grants permissions to a user to access his own SageMaker notebook instance. Attach the policy to each member's IAM user.

Create an AWS KMS Customer Master Key (CMK) for each member. Instruct the members to use their personal KMS CMK to encrypt their training data in Amazon S3.

Create a VPC for every member and use AWS PrivateLink to associate one interface endpoint per VPC.

QUESTION 5

Topic: Machine Learning Implementation and Operations

A company uses Amazon DynamoDB to store user-interaction data generated from a website. A Machine Learning Specialist was instructed to use this data in designing a product recommendation system.

Which approach will enable the Specialist to use Amazon SageMaker to train a model?

Use AWS DMS to migrate the DynamoDB table data into a MySQL database hosted on Amazon RDS. Provide the new database endpoint within the notebook instance.

Export the DynamoDB table data into an Amazon S3 bucket using AWS Data Pipeline and provide the S3 path of the data within the notebook instance.

Enable DynamoDB Streams to copy the data into a new table and provide the DynamoDB Stream endpoint within the notebook instance.

Pull the data from the database by configuring a direct connection between the notebook instance and the DynamoDB table.

QUESTION 6

Topic: Exploratory Data Analysis

A Machine Learning Specialist is preparing the dataset to be used for training a linear learner model in Amazon SageMaker. During exploratory data analysis, he has detected multiple feature columns that have missing values. The percentage of missing data across the whole training dataset is about 10%. The Specialist is worried that this might cause bias to his model that can lead to inaccurate results.

Which approach will MOST likely yield the best result in reducing the bias caused by missing values?

Use supervised learning methods to estimate the missing values for each feature.

Compute the mean of non-missing values in the same column and use the result to replace missing values.

Drop the columns that include missing values because they only account for 10% of the training data.

Compute the mean of non-missing values in the same row and use the result to replace missing values.

An on-premises application ingests hundreds of English audio recordings that are sent to Amazon S3 every day. A Machine Learning Specialist must develop a solution for reviewing their conversation topics.

Which approach will allow the Specialist to complete the task in the LEAST amount of effort?

Execute an Amazon Transcribe job for each audio file located on Amazon S3. Then, analyze the result with Amazon Comprehend.

Execute an Amazon Transcribe job for each audio file located on Amazon S3. Then, analyze the result using an NLP model trained with Amazon SageMaker.

Transcribe the audio files using the Amazon SageMaker BlazingText algorithm. Then, analyze the result with Amazon Transcribe.

Execute an Amazon Translate job for each audio file located on Amazon S3. Then, analyze the result with Amazon Comprehend.

Transcriber
Topic

SAGE

QUESTION 8

Topic: Exploratory Data Analysis

A Machine Learning Specialist has created a model without performing an initial investigation on the training data. The training time is slow and the predictive ability of the model is unstable during inference. A co-worker recommended that he should run the Principal Component Analysis (PCA) algorithm on the data before starting to train the model.

Which is a valid effect of the said algorithm?

PCA will reduce the dimensionality of the data based on correlations.

PCA will attempt to increase the input data available for the model without relying on new data.

PCA will transform category variables into vectorized numerical data.

PCA will impute missing data.

QUESTION 9

Topic: Modeling

A Machine Learning Specialist was assigned to develop a model for solving a customer segmentation task that will determine the needed social program by state or by city. The responses from the collection of 300 questions, answered by each citizen in various cities, will be used as the training data.

Which combination of algorithms is the MOST fitting to solve this task? (Select TWO.)

Factorization machines algorithm

K-nearest neighbor algorithm

K-means algorithm

Logistic Regression algorithm

Principal Component Analysis (PCA) algorithm

[Home](#)[Certifications](#)[Free Test](#)[Blog](#)

MACHINE LEARNING HAS BECOME POPULAR AMONG IT ENTHUSIASTS. THE

AWS Machine Learning specialty exam is intended to cover Amazon Web Services products that enable developers to discover patterns in end-user data using algorithms, build mathematical models based on these patterns, and then create and implement predictive applications. Every organization wants its most valuable asset – its workforce – to be technologically up to date at all times.

And, if you work for an IT company, staying current on technological developments is critical to maintaining your position and reputation. Certifications also demonstrate your dedication to your work and your organization. Keeping yourself up to date boosts your confidence and allows you to stand out in a crowd.

About the Certification

The AWS Certified Machine Learning – Specialty (MLS-C01) exam is designed for people who work in development or data science. This exam validates an examinee's ability to use the AWS Cloud to build, train, tune, and deploy machine learning (ML) models.

It assesses an examinee's ability to design, implement, deploy, and maintain machine learning solutions for specific business



100%
Exam
Pass
Guarantee

Want to
qualify AWS
machine
learning
specialty
exam?

Learn
More

Are you
preparing for
AWS
machine

problems. It will confirm the candidate's capability to:

- Choose and justify the best ML approach for a given business problem.
- Determine the best AWS services for implementing ML solutions.
- Create and implement scalable, cost-effective, dependable, and secure machine learning solutions.

learning
specialty
exam?

Take
a
Quiz

AWS Certified Machine Learning – Specialty Sample Questions

Question 1

While performing mini-batch training on a neural network for a classification task, a Data Scientist notices oscillations in training accuracy. Which of the following is MOST LIKELY to be the CAUSE of this problem?

- A. The dataset's class distribution is skewed.
- B. Dataset shuffle is turned off.
- C. The batch size is insufficient.
- D. The learning rate is exceptionally rapid.

Correct Answer – D

Reference: <https://towardsdatascience.com/deep-learning-personal-notes-part-1-lesson-2-8946fe970b95>

Question 2

When submitting Amazon SageMaker training tasks that use one of the built-in algorithms, which common parameters MUST be provided? (Choose three.)

- A. The training channel identifies training data on an Amazon S3 bucket.
- B. The validation channel identifies the validation data's location on an Amazon S3 bucket.
- C. The IAM role that Amazon SageMaker can use to perform tasks on the users' behalf.
- D. Hyperparameters in a JSON array, as specified by the algorithm.
- E. The Amazon EC2 instance class indicates whether training will be performed with a CPU or a GPU.
- F. The output path specifies where the trained model will be stored on an Amazon S3 bucket.

Correct Answer - A,E,F

Question 3

A retail chain has been using Amazon Kinesis Data Firehose to load purchase data from its 20,000-store network into Amazon S3. Training data will require additional but simple transformations, and certain characteristics will be merged, to facilitate the training of a more advanced machine learning model. The model must be retrained on a daily basis.

Given a large number of stores and historical data ingestion, which update will require the LEAST amount of development work?

- A. Require stores to capture their data locally on AWS

Storage Gateway for loading into Amazon S3, then use AWS Glue to transform it.

- B. Deploy an Amazon EMR cluster with Apache Spark and the transformation logic, and have the cluster run every day on the accumulating records in Amazon S3, outputting new/transformed records to Amazon S3.
- C. Launch a fleet of Amazon EC2 instances containing the transformation logic, instruct them to transform the data records accumulating on Amazon S3, and then output the transformed records to Amazon S3.
- D. Add an Amazon Kinesis Data Analytics stream downstream of the Kinesis Data Firehose stream that uses SQL to transform raw record attributes into simple transformed values.

Correct Answer - D

Question 4

A data scientist uses an Amazon SageMaker notebook instance to explore and analyze data. This entails installing on the notebook instance some Python packages that are not natively available on Amazon SageMaker. How can a machine learning expert ensure that the data scientist's essential packages are always available on the notebook instance?

- A. Set up the AWS Systems Manager Agent on the underlying Amazon EC2 instance and use Systems Manager Automation to run the package installation commands.
- B. Create a Jupyter notebook file (.ipynb) with cells containing the package installation commands to run and save it to the /etc/init directory of each Amazon SageMaker

notebook instance.

- C. From the Jupyter notebook console, use the conda package manager to apply the necessary conda packages to the notebook's default kernel.
- D. Using the package installation commands, create an Amazon SageMaker lifecycle configuration and assign it to the notebook instance.

Correct Answer - B

Reference: <https://towardsdatascience.com/automating-aws-sagemaker-notebooks-2dec62bc2c84>

Question 5

A web-based company wants to boost conversions on its landing page. Using a large historical dataset of client visits, the company developed a multi-class deep learning network algorithm using Amazon SageMaker on a regular basis. However, there is an overfitting problem: training data shows a prediction accuracy of 90%, whereas test data shows a prediction accuracy of 70%. In order to optimize visit-to-purchase conversions, the organization must improve the generalizability of its model before putting it into production.

TOO COMPLEX

Which activity is recommended to ensure that the company's test and validation data is modeled as accurately as possible?

- A. Increase the randomization of training data in training mini-batches.
- B. Assign a larger proportion of the total data to the training dataset.

- C. Incorporate L1 or L2 regularisation as well as dropouts into the training.
- D. Reduce the number of deep learning network layers and units (or neurones).

Correct Answer - D

Question 6

A long short-term memory (LSTM) model is used by a company to evaluate the risk variables associated with a specific energy sector. The programme analyses multi-page text documents and classifies each phrase as either dangerous or not dangerous. Despite the Data Scientist's extensive experimentation with various network architectures and tuning of the associated hyperparameters, the model underperforms. Which technique will result in the MAXIMUM performance increase?

Training another
model does not
guarantee
anything

- A. Pretrain term frequency-inverse document frequency (TF-IDF) vectors on a large collection of energy-related news articles to initialise the words.
- B. Instead of LSTM, use gated recurrent units (GRUs) and run the training process until the validation loss stops decreasing.
- C. Lower the learning rate and repeat the training process until the training loss no longer decreases.
- D. Pretrain the words using word2vec embeddings on a large collection of energy-related news articles.

Correct Answer - C

Question 7

An eCommerce startup is using photographs to automate product categorization. A data scientist trained a computer vision model using the Amazon SageMaker image categorization method. The photos for each product are organized by the product line. The model's accuracy is insufficient when classifying new items. All product photos are the same size and are stored in an Amazon S3 bucket. The company wants to improve the model as soon as possible so that it can be used for future products. Which actions would improve the accuracy of the solution? (Choose three.)

- A. To improve accuracy, use the SageMaker semantic segmentation algorithm to train a new model.
- B. Classify the products in the dataset using the Amazon Rekognition DetectLabels API.
- C. Enhance the dataset's images. Crop, resize, flip, rotate, and adjust the brightness and contrast of the images using open source libraries.
- D. Use a SageMaker notebook to implement pixel normalization and image scaling. Amazon S3 should be used to store the new dataset.
- E. Train a new model with Amazon Rekognition Custom Labels.
- F. Examine the product categories for class imbalances and use oversampling or undersampling as needed. Amazon S3 should be used to store the new dataset.

Correct Answer – B, C, E

Reference:

<https://docs.aws.amazon.com/rekognition/latest/dg/how-it-works.html>

[types.html https://towardsdatascience.com/image-processing-techniques-for-computer-vision-11f92f511e21](https://towardsdatascience.com/image-processing-techniques-for-computer-vision-11f92f511e21)

<https://docs.aws.amazon.com/rekognition/latest/customlabels-dg/training-model.html>

Question 8

An employee saw a video clip with audio on a company's social media page. The video is only available in Spanish. The employee's first language is English, and he or she cannot understand Spanish. Sentiment analysis is requested by the employee. Which service combination is the MOST EFFECTIVE in terms of task completion?

audio → text
Transcribe
translate
→ comprehend

- A. Amazon Transcribe, Amazon Translate, and Amazon Comprehend are three services provided by Amazon.
- ✗ B. Amazon Transcribe, Amazon Comprehend, and Amazon SageMaker seq2seq are all available.
- ✗ C. Amazon Transcribe, Amazon Translate, and Amazon SageMaker Neural Topic Model are examples of such services (NTM)
- ✗ D. Amazon Transcribe, Amazon Translate, and Amazon SageMaker BlazingText are all available on Amazon.

Correct Answer - A

Question 9

A company examines camera photos of the tops of objects placed on store shelves to determine which items have been taken and which remain. After many hours of data tagging, the organization now has a total of 1,000 hand-labeled photos encompassing ten distinct things. The instruction was ineffective. Which machine

learning technique best meets the company's long-term objectives?

- A. Grayscale the images and retrain the model
- B. Reduce the number of distinct items from ten to two, construct the model, and iterate.
- C. Attach different coloured labels to each item, take new photos, and reassemble the model.
- D. Use image variants such as inversions and translations to augment training data for each item, then build the model and iterate.

Correct Answer - A

Question 10

A plane engine manufacturer is compiling a time series of 200 performance indicators. During testing, engineers require near-real-time detection of significant production problems. All data must be saved for later analysis. Which strategy would be the MOST EFFECTIVE in terms of near-real-time defect detection?

- A. Make use of AWS IoT Analytics for data ingestion, storage, and analysis. To perform anomaly analysis, use Jupyter notebooks from within AWS IoT Analytics.
- B. Use Amazon S3 for data ingestion, storage, and analysis. To detect anomalies, use an Amazon EMR cluster to perform Apache Spark ML k-means clustering.
- C. Use Amazon S3 for data ingestion, storage, and analysis. To detect anomalies, use the Amazon SageMaker Random Cut Forest (RCF) algorithm.

- D. Perform anomaly detection using Amazon Kinesis Data Firehose and Amazon Kinesis Data Analytics Random Cut Forest (RCF). Kinesis Data Firehose can be used to store data in Amazon S3 for later analysis.

→ Real time
use firehose, kinesis
cut anomaly
detection

Correct Answer - B

Question 11

Amazon A company is using Polly to convert plaintext texts to voice in order to automate corporate announcements. However, corporate acronyms are mispronounced in modern papers. What should a Machine Learning Specialist do about this problem in the future?

- A. Convert existing documents to SSML and add pronunciation tags.
- B. Create a lexicon of appropriate pronunciations.
- C. Use speech marks to help with pronunciation.
- D. Preprocess the text files for pronunciation using Amazon Lex.

→ S →

Correct Answer - A

Reference: <https://docs.aws.amazon.com/polly/latest/dg/ssml.html>

Question 12

A manufacturing company asks a machine learning expert for help in developing a model that categorizes damaged components into one of eight defect classes. The company provided over 100,000 photos per fault category for training purposes. According to the

expert, the validation accuracy of the picture classification model is 80%, while the training accuracy is 90%. Human-level performance for this type of image categorization is estimated to be around 90%. What should the professional take into account when resolving this situation?

- A. More training time
- B. Extending the network
- C. Making use of a different optimizer
- D. Making use of some form of regularisation

Correct Answer - D

Reference: <https://acloud.guru/forums/aws-certified-machine-learning-specialty/discussion/-MGdBUKmQ02zC3uOq4VL/AWS%20Exam%20Machine%20Learning>

Question 13

A company is creating a demand forecasting model based on machine learning (ML). During the development stage, an ML expert performs feature engineering on an Amazon SageMaker laptop with limited CPU and memory resources. A data engineer typically uses the same notebook to perform data preparation once a day, which requires a large amount of RAM and takes only two hours to complete. The data preparation is not intended to benefit from GPU acceleration. All processes on an ml.m5.4xlarge notebook instance are running normally.

The organization receives an AWS Budgets warning that the billing for this month exceeds the budgeted amount.

Which of the following solutions will save you the most money?

- A. Set the notebook instance type to memory optimized with the same vCPU number as the ml.m5.4xlarge instance. When not in use, turn off the notebook. In that instance, run both data preprocessing and feature engineering development.
- B. Maintain the same notebook instance type and size. When not in use, turn off the notebook. Using Amazon SageMaker Processing, run data preprocessing on a P3 instance type with the same memory as the ml.m5.4xlarge instance.
- C. Switch to a smaller general-purpose notebook instance. When not in use, turn off the notebook. Using Amazon SageMaker Processing, run data preprocessing on an ml.r5 instance with the same memory size as the ml.m5.4xlarge instance.
- D. Replace the notebook instance with a smaller general-purpose instance. When not in use, turn off the notebook. Using the Reserved Instance option, run data preprocessing on an R5 instance with the same memory size as the ml.m5.4xlarge instance.

Correct Answer - B

Question 14

Telemetry data is generated by wind turbines, weather stations, and solar panels for an energy company. The organization wishes to perform predictive maintenance on these devices. The devices are dispersed throughout the city and have spotty internet access. A team of data scientists is analyzing the telemetry data to detect anomalies and predict repairs before the devices fail. The team requires a data ingestion system that is scalable, secure, and capable of handling large amounts of data at high speeds. The team has decided to keep data on Amazon S3.

Which technique meets these requirements?

- A. Ingest the data by making an HTTP API call to an Amazon EC2 web server. To load the data into Amazon S3, set up EC2 instances in an Auto Scaling configuration behind an Elastic Load Balancer.
- B. Send the data to AWS IoT Core via Message Queuing Telemetry Transport (MQTT). Create an AWS IoT Core rule to send data to an Amazon Kinesis data stream configured to write to an S3 bucket using Amazon Kinesis Data Firehose.
- C. Send the data to AWS IoT Core via Message Queuing Telemetry Transport (MQTT). Create an AWS IoT Core rule to route all MQTT data to an Amazon Kinesis Data Firehose delivery stream configured to write to an S3 bucket.
- D. Message Queuing Telemetry Transport (MQTT) the data to an Amazon Kinesis data stream configured to write to an S3 bucket.

→ cannot write directly to S3

Correct Answer - C

Reference: <https://aws.amazon.com/blogs/industries/real-time-operational-monitoring-of-renewable-energy-assets-with-aws-iot/>

Question 15

A company wants to forecast home selling prices using historical sales data. In the company's dataset, the goal variable is the selling price. The attributes include the lot size, living space and non-living area measurements, the number of bedrooms and bathrooms, the year built, and the postal code. The company wants to forecast home sales prices using multivariable linear regression.

Which step should a machine learning expert take to remove extraneous data and simplify the model?

Outliers

- A. Create a histogram of the features and calculate the standard deviation. Remove features with a lot of variation.
- B. Create a histogram of the features and calculate the standard deviation. Features with low variance should be removed.
- C. Create a heatmap displaying the dataset's correlation with itself. Features with low mutual correlation scores should be removed.
- D. Perform a correlation analysis on all features in relation to the target variable. Features with low target variable correlation scores should be removed.

Correct Answer - D

Question 16

A Machine Learning Specialist collects customer data for an online shopping website. The data includes demographic information, previous visits, and information about the surrounding area. The Specialist is in charge of creating a machine learning strategy for identifying client purchasing habits, preferences, and trends in order to improve the website's service and recommendation capabilities.

What action should the Specialist recommend?

- A. Identifying patterns in the customer database using Latent Dirichlet Allocation (LDA) for the given collection of discrete data.

topic
modelling

- ~~B. A neural network with at least three layers and random initial weights to recognize patterns in the customer database.~~
- ~~C. Identifying patterns in the customer database through collaborative filtering based on user interactions and correlations.~~
- D. RCF over random subsamples to identify patterns in the customer database.

Correct Answer - C

Question 17

A mobile network operator is developing an analytics platform for analyzing and optimizing business operations using Amazon Athena and Amazon S3. The source systems send data in CSV format in real-time. The Data Engineering team wishes to convert the data to the Apache Parquet format before storing it on Amazon S3. Which approach necessitates the LEAST amount of effort to implement?

CSV → Parquet
6/6

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

Correct Answer - B

Question 18

A utility company wants to forecast future energy consumption for its residential and commercial clients. Data on historical energy consumption over the last decade is provided. A team of data scientists will conduct the initial data analysis and feature selection, which will include historical power use data as well as data on the weather, the number of people on the property, and public holidays. Data scientists use Amazon Forecast to generate the projections. Which Forecast algorithm should data scientists use to meet these criteria?

- A. AIM stands for Autoregressive Integrated Moving Average (AIRMA)
- B. Smoothing on the Exponential (ETS)
- C. Quantile Regression Using a Convolutional Neural Network (CNN-QR)
- D. Prophet

Correct Answer - B

Reference:

<https://jesit.springeropen.com/articles/10.1186/s43067-020-00021-8>

Question 19

A machine learning (ML) model is used by a retail company to forecast daily sales. The model has been producing incorrect results for the last three weeks, according to the company's brand

manager.

An AWS Glue task consolidates the forecasting input data with the actual daily sales data and the model's predictions at the end of each day. Using the AWS Glue task, the data is saved in Amazon S3.

Using an Amazon SageMaker Studio notebook, the company's machine learning team is analyzing the model's errors. What actions should the machine learning team take on the SageMaker Studio notebook to best demonstrate the model's degradation?

- A. Make a histogram of the last three weeks' daily sales. In addition, make a histogram of the daily sales prior to that time period.
- B. Produce a histogram of model errors over the last three weeks. In addition, make a histogram of the model errors prior to that time period.
- C. Make a line chart of the model's weekly mean absolute error (MAE).
- D. For the last three weeks, create a scatter plot of daily sales versus model error. Create a scatter plot of daily sales versus model error prior to that period.

Correct Answer - C

Reference: <https://machinelearningmastery.com/time-series-forecasting-performance-measures-with-python/>

Question 20

A mobile device manufacturer wishes to determine and adjust the optimal selling price for its products. The company is gathering and characterizing relevant data in order to train machine learning (ML) models. There are over 1,000 features, and the corporation

wants to know which ones are most important in determining the selling price. Which feature selection strategies should the company use? (Make three choices.)

- A. Scaling of data with standardization and normalization
- B. Heat map correlation plot
- C. Binning of data
- D. Univariate analysis
- E. Using a tree-based classifier to determine feature importance
- F. Data enhancement

Correct Answer – C, D, F

Reference: <https://towardsdatascience.com/an-overview-of-data-preprocessing-features-enrichment-automatic-feature-selection-60b0c12d75ad> <https://towardsdatascience.com/feature-selection-using-python-for-classification-problem-b5f00a1c7028#:~:text=Univariate%20feature%20selection%20works%>

Hurry up and start preparing now!

© 2020

TestPrepTraining



Disclaimer:

Testpretraining.com does not offer exam dumps or questions from actual exams. We offer learning material and practice tests created by subject matter experts to assist and help learners prepare for those exams. All certification brands used on the website are owned by the respective brand owners. Testpretraining does not own or claim any ownership on any of the brands.

QUESTION

A product owner is working with a data science team who use Spark on AWS EMR for big data projects and have significant expertise with Spark. They have indicated a preference to work with Spark + MLlib to address the real-time predictions use case you've proposed.

Which of the following would you recommend as the most scalable and efficient way to proceed?

- Use Spark for ETL, SageMaker to train & deploy models
- Build a Spark Pipeline to including ETL and model training. Deploy via the Spark cluster
- Use Spark for Model development, then deploy via SageMaker
- Use Spark for ETL and to train a model, MLEAP Serialisation to deploy into SageMaker

Next**QUESTION**

A product owner has made the development team aware of a number of built-in machine learning capabilities associated with various AWS Services. Match the machine learning capability with the AWS service.

- a) Anomaly detection (RCF) + Hotspot detection
 - b) Duplicate detection (FindMatches ML)
 - c) Anomaly detection (RCF), Forecasting, Auto-Narratives
 - d) Active label learning
- (Select One)

- a) Amazon QuickSight, b) AWS Glue, c) Kinesis Data Analytics d) SageMaker - Ground Truth

- a) AWS Glue b) Amazon QuickSight c) SageMaker - Ground Truth

- a) Kinesis Data Analytics b) EMR Pipeline c) Amazon QuickSight d) SageMaker - Ground Truth

- a) Kinesis Data Analytics , b) AWS Glue, c) Amazon QuickSight d) SageMaker - Ground Truth

QUESTION

The development team want to make use of SageMaker auto-scaling. Where should this be configured?

(Select One?)

- In the model, edit the instance type and add auto-scaling parameters

- In the endpoint configuration - add auto-scaling policy

- In the endpoint itself - configure the auto-scaling policy

- Configure SageMaker auto-scaling directly in CloudWatch

QUESTION

A data scientist is utilising SageMaker notebooks to explore various machine learning models with a subset of the data.

Which of the following are true?

(Select Two)

- Pipe mode should be used so that the notebook instance launches faster and training completes more quickly
- Use local mode where available to iterate and test work
- Local mode is available for all SageMaker estimators
- Local mode is available for Tensorflow and MXNet only

QUESTION

A data & development team have built a deep learning model on a SageMaker TensorFlow Framework container. However, the model is performing poorly and training is frequently failing.

Which of the following approaches would help resolve the issue?

(Select One)

- Manually print debug data and analyse that data
- use open source tools for charting and refer to CloudWatch logs
- add dropout layers and substitute tanh activation functions for relu
- define rules via Sagemaker debugger, create hooks to save tensors to S3, use SageMaker Studio to visualise training issues.

QUESTION

Which of the following must be set by the user (required hyperparameters), for SageMaker's built-in algorithm, Linear Learner? (assume classification)

(Select Two)

- feature_dim
- num_classes
- predictor_type
- epochs

QUESTION

A developer is choosing between two different approaches for working with SageMaker – which of the following are true?

(Select Two)

- SageMaker Python SDK provides finer grained control of SageMaker features than Boto3
- Boto3 is recommended for production and automation work with SageMaker
- SageMaker Python SDK provides several high-level abstractions for working with SageMaker including Estimators, Models, Predictors, Session, Transformers, Processors
- Unlike SageMaker Python SDK, Boto3 is available in languages other than Python, such as R

QUESTION

A data scientist has been experimenting with a deep neural network, utilising Kaggle's free Jupyter notebook environment, Keras and TensorFlow.

Which of the options corresponds to the minimal actions needed to get this running on SageMaker?

(Select One)

- a. Load train & test data into S3
- b. Modify your Kaggle Script to accept model directory, train and test, host arguments and to save the model
- c. Add your Kaggle script (myscript.py) into your Notebook instance
- d. Register Your Container on ECM
- e. Specify the Container location in the SageMaker estimator.
- f. Create a SageMaker notebook and use myscript.py as the entry point in the TensorFlow estimator.
- g. Use SageMaker Kaggle Sync to transfer the script file and model artifacts to S3.
- h. Create a serve script

- a, e, f, h

- h, b, d, e

- a, c, d, h

- a,b,c,f

☰

a deep neural network, utilising Kaggle's free Jupyter notebook environment, Keras and Tensorflow.
Which of the options corresponds to the minimal actions needed to get this running on SageMaker?
(Select One)

a. Load train & test data into S3
b. Modify your Kaggle Script to accept model directory, train and test, host arguments and to save the model
c. Add your Kaggle script (myscript.py) into your Notebook instance
d. Register Your Container on ECM
e. Specify the Container location in the SageMaker estimator.
f. Create a SageMaker notebook and use myscript.py as the entry point in the Tensorflow estimator.
g. Use SageMaker Kaggle Sync to transfer the script file and model artifacts to S3.
h. Create a serve script

a, e, f, h
 h, b, d, e
 a, c, d, h
 a,b,c,f

[Prev](#) [Next](#)

#aws-certification

aws-certification