DUMPSARENA

AWS Certified Machine Learning - Specialty

Amazon AWS AWS-Certified-Machine-Learning-Specialty-MLS-C01

Version Demo

Total Demo Questions: 15

Total Premium Questions: 377

Buy Premium PDF

https://dumpsarena.co

sales@dumpsarena.co

sales@dumpsarena.co
dumpsarena.co

QUESTION NO: 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.

ANSWER: C D

QUESTION NO: 2

A machine learning (ML) specialist is using Amazon SageMaker hyperparameter optimization (HPO) to improve a model's accuracy. The learning rate parameter is specified in the following HPO configuration:

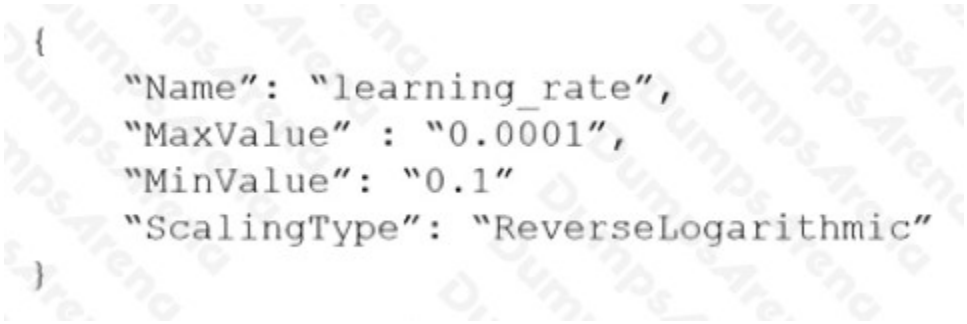
```
{
    "Name": "learning_rate",
    "MaxValue" : "0.0001",
    "MinValue": "0.1"
}
```

During the results analysis, the ML specialist determines that most of the training jobs had a learning rate between 0.01 and 0.1. The best result had a learning rate of less than 0.01. Training jobs need to run regularly over a changing dataset. The ML specialist needs to find a tuning mechanism that uses different learning rates more evenly from the provided range between MinValue and MaxValue.

Which solution provides the MOST accurate result?

A-

Modify the HPO configuration as follows:



Select the most accurate hyperparameter configuration form this HPO job.

B-

Run three different HPO jobs that use different learning rates form the following intervals for MinValue and MaxValue while using the same number of training jobs for each HPO job:

[0.01, 0.1]

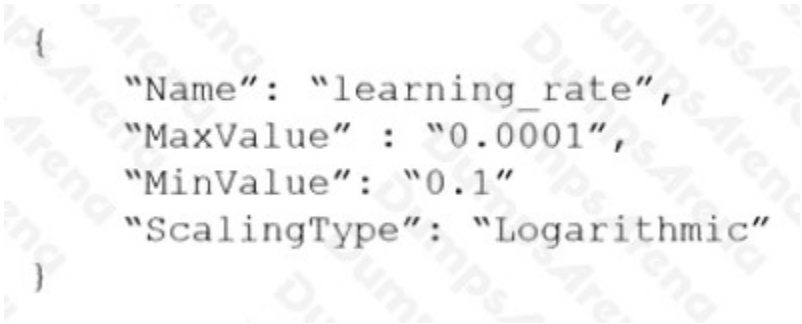
[0.001, 0.01]

[0.0001, 0.001]

Select the most accurate hyperparameter configuration form these three HPO jobs.

C-

Modify the HPO configuration as follows:



Select the most accurate hyperparameter configuration form this training job.

D-

Run three different HPO jobs that use different learning rates form the following intervals for MinValue and MaxValue. Divide the number of training jobs for each HPO job by three:

[0.01, 0.1]

[0.001, 0.01]

[0.0001, 0.001]

Select the most accurate hyperparameter configuration form these three HPO jobs.

A. Option A

B. Option B

C. Option C

D. Option D

ANSWER: C

QUESTION NO: 3

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.

ANSWER: C D

QUESTION NO: 4

A company is building a predictive maintenance model based on machine learning (ML). The data is stored in a fully private Amazon S3 bucket that is encrypted at rest with AWS Key Management Service (AWS KMS) CMKs. An ML specialist must run data preprocessing by using an Amazon SageMaker Processing job that is triggered from code in an Amazon SageMaker notebook. The job should read data from Amazon S3, process it, and upload it back to the same S3 bucket. The preprocessing code is stored in a container image in Amazon Elastic Container Registry (Amazon ECR). The ML specialist needs to grant permissions to ensure a smooth data preprocessing workflow.

Which set of actions should the ML specialist take to meet these requirements?

A. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs, S3 read and write access to the relevant S3 bucket, and appropriate KMS and ECR permissions. Attach the role to the SageMaker notebook instance. Create an Amazon SageMaker Processing job from the notebook.

B. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs. Attach the role to the SageMaker notebook instance. Create an Amazon SageMaker Processing job with an IAM role that has read and write permissions to the relevant S3 bucket, and appropriate KMS and ECR permissions.

C. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs and to access Amazon ECR. Attach the role to the SageMaker notebook instance. Set up both an S3 endpoint and a KMS endpoint in the default VPCreate Amazon SageMaker Processing jobs from the notebook.

D. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs. Attach the role to the SageMaker notebook instance. Set up an S3 endpoint in the default VPC. Create Amazon SageMaker Processing jobs with the access key and secret key of the IAM user with appropriate KMS and ECR permissions.

ANSWER: D

QUESTION NO: 5

The chief editor for a product catalog wants the research and development team to build a machine learning system that can be used to detect whether or not individuals in a collection of images are wearing the company's retail brand. The team has a set of training data.

Which machine learning algorithm should the researchers use that BEST meets their requirements?

A. Latent Dirichlet Allocation (LDA)

B. Recurrent neural network (RNN)

C. K-means

D. Convolutional neural network (CNN)

ANSWER: D

QUESTION NO: 6

A Data Scientist is building a model to predict customer churn using a dataset of 100 continuous numerical features. The Marketing team has not provided any insight about which features are relevant for churn prediction. The Marketing team wants to interpret the model and see the direct impact of relevant features on the model outcome. While training a logistic regression model, the Data Scientist observes that there is a wide gap between the training and validation set accuracy.

Which methods can the Data Scientist use to improve the model performance and satisfy the Marketing team’s needs? (Choose two.)

A. Add L1 regularization to the classifier

B. Add features to the dataset

C. Perform recursive feature elimination

D. Perform t-distributed stochastic neighbor embedding (t-SNE)

E. Perform linear discriminant analysis

ANSWER: B E

QUESTION NO: 7

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.

F. The output path specifying where on an Amazon S3 bucket the trained model will persist.

ANSWER: A E F

QUESTION NO: 8

A library is developing an automatic book-borrowing system that uses Amazon Rekognition. Images of library members' faces are stored in an Amazon S3 bucket. When members borrow books, the Amazon Rekognition CompareFaces API operation compares real faces against the stored faces in Amazon S3.

The library needs to improve security by making sure that images are encrypted at rest. Also, when the images are used with Amazon Rekognition. they need to be encrypted in transit. The library also must ensure that the images are not used to improve Amazon Rekognition as a service.

How should a machine learning specialist architect the solution to satisfy these requirements?

A. Enable server-side encryption on the S3 bucket. Submit an AWS Support ticket to opt out of allowing images to be used for improving the service, and follow the process provided by AWS Support.

B. Switch to using an Amazon Rekognition collection to store the images. Use the IndexFaces and SearchFacesByImage API operations instead of the CompareFaces API operation.

C. Switch to using the AWS GovCloud (US) Region for Amazon S3 to store images and for Amazon Rekognition to compare faces. Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.

D. Enable client-side encryption on the S3 bucket. Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.

ANSWER: B

QUESTION NO: 9

A company supplies wholesale clothing to thousands of retail stores. A data scientist must create a model that predicts the daily sales volume for each item for each store. The data scientist discovers that more than half of the stores have been in business for less than 6 months. Sales data is highly consistent from week to week. Daily data from the database has been aggregated weekly, and weeks with no sales are omitted from the current dataset. Five years (100 MB) of sales data is available in Amazon S3.

Which factors will adversely impact the performance of the forecast model to be developed, and which actions should the data scientist take to mitigate them? (Choose two.)

A. Detecting seasonality for the majority of stores will be an issue. Request categorical data to relate new stores with similar stores that have more historical data.

B. The sales data does not have enough variance. Request external sales data from other industries to improve the model's ability to generalize.

C. Sales data is aggregated by week. Request daily sales data from the source database to enable building a daily model.

D. The sales data is missing zero entries for item sales. Request that item sales data from the source database include zero entries to enable building the model.

E. Only 100 MB of sales data is available in Amazon S3. Request 10 years of sales data, which would provide 200 MB of training data for the model.

ANSWER: A B

Explanation:

Reference: https://towardsdatascience.com/sales-forecasting-from-time-series-to-deep-learning-5d115514bfac
https://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf

QUESTION NO: 10

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

Which combination of algorithms would provide the appropriate insights? (Choose two.)

A. The factorization machines (FM) algorithm

B. The Latent Dirichlet Allocation (LDA) algorithm

C. The principal component analysis (PCA) algorithm

D. The k-means algorithm

E. The Random Cut Forest (RCF) algorithm

ANSWER: C D

Explanation:

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

QUESTION NO: 11

A data scientist is using an Amazon SageMaker notebook instance and needs to securely access data stored in a specific Amazon S3 bucket.

How should the data scientist accomplish this?

A. Add an S3 bucket policy allowing GetObject, PutObject, and ListBucket permissions to the Amazon SageMaker notebook ARN as principal.

B. Encrypt the objects in the S3 bucket with a custom AWS Key Management Service (AWS KMS) key that only the notebook owner has access to.

C. Attach the policy to the IAM role associated with the notebook that allows GetObject, PutObject, and ListBucket operations to the specific S3 bucket.

D. Use a script in a lifecycle configuration to configure the AWS CLI on the instance with an access key ID and secret.

ANSWER: C

QUESTION NO: 12

A company wants to predict the sale prices of houses based on available historical sales data. The target variable in the company's dataset is the sale price. The features include parameters such as the lot size, living area measurements, non-living area measurements, number of bedrooms, number of bathrooms, year built, and postal code. The company wants to use multi-variable linear regression to predict house sale prices.

Which step should a machine learning specialist take to remove features that are irrelevant for the analysis and reduce the model's complexity?

A. Plot a histogram of the features and compute their standard deviation. Remove features with high variance.

B. Plot a histogram of the features and compute their standard deviation. Remove features with low variance.

C. Build a heatmap showing the correlation of the dataset against itself. Remove features with low mutual correlation scores.

D. Run a correlation check of all features against the target variable. Remove features with low target variable correlation scores.

ANSWER: D

QUESTION NO: 13

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

Which of the following services can feed data to the MapReduce jobs? (Choose two.)

A. AWS DMS

B. Amazon Kinesis

C. AWS Data Pipeline

D. Amazon Athena

E. Amazon ES

ANSWER: A E

QUESTION NO: 14

An agricultural company is interested in using machine learning to detect specific types of weeds in a 100-acre

grassland field. Currently, the company uses tractor-mounted cameras to capture multiple images of the field

as 10 × 10 grids. The company also has a large training dataset that consists of annotated images of popular

weed classes like broadleaf and non-broadleaf docks.

The company wants to build a weed detection model that will detect specific types of weeds and the location of each type within the field. Once the model is ready, it will be hosted on Amazon SageMaker endpoints. The

model will perform real-time inferencing using the images captured by the cameras.

Which approach should a Machine Learning Specialist take to obtain accurate predictions?

A. Prepare the images in RecordIO format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an image classification algorithm to categorize images into various weed classes.

B. Prepare the images in Apache Parquet format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an object-detection single-shot multibox detector (SSD) algorithm.

C. Prepare the images in RecordIO format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an object-detection single-shot multibox detector (SSD) algorithm.

D. Prepare the images in Apache Parquet format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an image classification algorithm to categorize images into various weed classes.

ANSWER: C

QUESTION NO: 15

A company is building a line-counting application for use in a quick-service restaurant. The company wants to use video cameras pointed at the line of customers at a given register to measure how many people are in line and deliver notifications to managers if the line grows too long. The restaurant locations have limited bandwidth for connections to external services and cannot accommodate multiple video streams without impacting other operations.

Which solution should a machine learning specialist implement to meet these requirements?

A. Install cameras compatible with Amazon Kinesis Video Streams to stream the data to AWS over the restaurant's existing internet connection. Write an AWS Lambda function to take an image and send it to Amazon Rekognition to count the number of faces in the image. Send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

B. Deploy AWS DeepLens cameras in the restaurant to capture video. Enable Amazon Rekognition on the AWS DeepLens device, and use it to trigger a local AWS Lambda function when a person is recognized. Use the Lambda function to send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

C. Build a custom model in Amazon SageMaker to recognize the number of people in an image. Install cameras compatible with Amazon Kinesis Video Streams in the restaurant. Write an AWS Lambda function to take an image. Use the SageMaker endpoint to call the model to count people. Send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

D. Build a custom model in Amazon SageMaker to recognize the number of people in an image. Deploy AWS DeepLens cameras in the restaurant. Deploy the model to the cameras. Deploy an AWS Lambda function to the cameras to use the model to count people and send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

ANSWER: A