Free Questions for MLS-C01 by certscare

Shared by Battle on 06-06-2022

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

A real-estate company is launching a new product that predicts the prices of new houses. The historical data for the properties and prices is stored in .csv format in an Amazon S3 bucket. The data has a header, some categorical fields, and some missing values. The company's data scientists have used Python with a common open-source library to fill the missing values with zeros. The data scientists have dropped all of the categorical fields and have trained a model by using the open-source linear regression algorithm with the default parameters.

The accuracy of the predictions with the current model is below 50%. The company wants to improve the model performance and launch the new product as soon as possible.

Which solution will meet these requirements with the LEAST operational overhead?

Options:

A- Create a service-linked role for Amazon Elastic Container Service (Amazon ECS) with access to the S3 bucket. Create an ECS cluster that is based on an AWS Deep Learning Containers image. Write the code to perform the feature engineering. Train a logistic regression model for predicting the price, pointing to the bucket with the dataset. Wait for the training job to complete. Perform the inferences.

B- Create an Amazon SageMaker notebook with a new IAM role that is associated with the notebook. Pull the dataset from the S3 bucket. Explore different combinations of feature engineering transformations, regression algorithms, and hyperparameters. Compare all the results in the notebook, and deploy the most accurate configuration in an endpoint for predictions.

C- Create an IAM role with access to Amazon S3, Amazon SageMaker, and AWS Lambda. Create a training job with the SageMaker built-in XGBoost model pointing to the bucket with the dataset. Specify the price as the target feature. Wait for the job to complete. Load the model artifact to a Lambda function for inference on prices of new houses.

D- Create an IAM role for Amazon SageMaker with access to the S3 bucket. Create a SageMaker AutoML job with SageMaker Autopilot pointing to the bucket with the dataset. Specify the price as the target attribute. Wait for the job to complete. Deploy the best model for predictions.

Answer:

A

Question 2

Question Type: MultipleChoice

A company has an ecommerce website with a product recommendation engine built in TensorFlow. The recommendation engine endpoint is hosted by Amazon SageMaker. Three compute-optimized instances support the expected peak load of the website.

Response times on the product recommendation page are increasing at the beginning of each month. Some users are encountering errors. The website receives the majority of its traffic between 8 AM and 6 PM on weekdays in a single time zone.

Which of the following options are the MOST effective in solving the issue while keeping costs to a minimum? (Choose two.)

Options:

A- Configure the endpoint to use Amazon Elastic Inference (EI) accelerators.

B- Create a new endpoint configuration with two production variants.

C- Configure the endpoint to automatically scale with the InvocationsPerInstance metric.

D- Deploy a second instance pool to support a blue/green deployment of models.

E- Reconfigure the endpoint to use burstable instances.

Answer:

B, D

Explanation:

https://www.redhat.com/en/topics/devops/what-is-blue-green-deployment

Question 3

Question Type: MultipleChoice

A manufacturing company wants to use machine learning (ML) to automate quality control in its facilities. The facilities are in remote locations and have limited internet connectivity. The company has 20 of training data that consists of labeled images of defective product parts. The training data is in the corporate on-premises data center.

The company will use this data to train a model for real-time defect detection in new parts as the parts move on a conveyor belt in the facilities. The company needs a solution that minimizes costs for compute infrastructure and that maximizes the scalability of resources for training. The solution also must facilitate the company's use of an ML model in the low-connectivity environments.

Which solution will meet these requirements?

Options:

A- Move the training data to an Amazon S3 bucket. Train and evaluate the model by using Amazon SageMaker. Optimize the model by using SageMaker Neo. Deploy the model on a SageMaker hosting services endpoint.

B- Train and evaluate the model on premises. Upload the model to an Amazon S3 bucket. Deploy the model on an Amazon SageMaker hosting services endpoint.

C- Move the training data to an Amazon S3 bucket. Train and evaluate the model by using Amazon SageMaker. Optimize the model by using SageMaker Neo. Set up an edge device in the manufacturing facilities with AWS IoT Greengrass. Deploy the model on the edge device.

D- Train the model on premises. Upload the model to an Amazon S3 bucket. Set up an edge device in the manufacturing facilities with AWS IoT Greengrass. Deploy the model on the edge device.

Answer:

A

Question 4

Question Type: MultipleChoice

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?

Options:

A- Modify the HPO configuration as follows:

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

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:

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

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.

Answer:

C

Question 5

Question Type: MultipleChoice

A retail company wants to update its customer support system. The company wants to implement automatic routing of customer claims to different queues to prioritize the claims by category.

Currently, an operator manually performs the category assignment and routing. After the operator classifies and routes the claim, the company stores the claim's record in a central database. The claim's record includes the claim's category.

The company has no data science team or experience in the field of machine learning (ML). The company's small development team needs a solution that requires no ML expertise.

Which solution meets these requirements?

Options:

A- Export the database to a .csv file with two columns: claim_label and claim_text. Use the Amazon SageMaker Object2Vec algorithm and the .csv file to train a model. Use SageMaker to deploy the model to an inference endpoint. Develop a service in the application to use the inference endpoint to process incoming claims, predict the labels, and route the claims to the appropriate queue.

B- Export the database to a .csv file with one column: claim_text. Use the Amazon SageMaker Latent Dirichlet Allocation (LDA) algorithm and the .csv file to train a model. Use the LDA algorithm to detect labels automatically. Use SageMaker to deploy the model to an inference endpoint. Develop a service in the application to use the inference endpoint to process incoming claims, predict the labels, and route the claims to the appropriate queue.

C- Use Amazon Textract to process the database and automatically detect two columns: claim_label and claim_text. Use Amazon Comprehend custom classification and the extracted information to train the custom classifier. Develop a service in the application to use the Amazon Comprehend API to process incoming claims, predict the labels, and route the claims to the appropriate queue.

D- Export the database to a .csv file with two columns: claim_label and claim_text. Use Amazon Comprehend custom classification and the .csv file to train the custom classifier. Develop a service in the application to use the Amazon Comprehend API to process incoming claims, predict the labels, and route the claims to the appropriate queue.

Answer:

C

Question 6

Question Type: MultipleChoice

A machine learning (ML) specialist needs to extract embedding vectors from a text series. The goal is to provide a ready-to-ingest feature space for a data scientist to develop downstream ML predictive models. The text consists of curated sentences in English. Many sentences use similar words but in different contexts. There are questions and answers among the sentences, and the embedding space must differentiate between them.

Which options can produce the required embedding vectors that capture word context and sequential QA information? (Choose two.)

Options:

A- Amazon SageMaker seq2seq algorithm

B- Amazon SageMaker BlazingText algorithm in Skip-gram mode

C- Amazon SageMaker Object2Vec algorithm

D- Amazon SageMaker BlazingText algorithm in continuous bag-of-words (CBOW) mode

E- Combination of the Amazon SageMaker BlazingText algorithm in Batch Skip-gram mode with a custom recurrent neural network (RNN)

Answer:

A, C

Explanation:

https://docs.aws.amazon.com/sagemaker/latest/dg/object2vec.html

Question 7

Question Type: MultipleChoice

A global financial company is using machine learning to automate its loan approval process. The company has a dataset of customer information. The dataset contains some categorical fields, such as customer location by city and housing status. The dataset also includes financial fields in different units, such as account balances in US dollars and monthly interest in US cents.

The company's data scientists are using a gradient boosting regression model to infer the credit score for each customer. The model has a training accuracy of 99% and a testing accuracy of 75%. The data scientists want to improve the model's testing accuracy.

Which process will improve the testing accuracy the MOST?

Options:

A- Use a one-hot encoder for the categorical fields in the dataset. Perform standardization on the financial fields in the dataset. Apply L1 regularization to the data.

B- Use tokenization of the categorical fields in the dataset. Perform binning on the financial fields in the dataset. Remove the outliers in

the data by using the z-score.

C- Use a label encoder for the categorical fields in the dataset. Perform L1 regularization on the financial fields in the dataset. Apply L2 regularization to the data.

D- Use a logarithm transformation on the categorical fields in the dataset. Perform binning on the financial fields in the dataset. Use imputation to populate missing values in the dataset.

Answer:

B

Question 8

Question Type: **MultipleChoice**

An ecommerce company sends a weekly email newsletter to all of its customers. Management has hired a team of writers to create additional targeted content. A data scientist needs to identify five customer segments based on age, income, and location. The customers' current segmentation is unknown. The data scientist previously built an XGBoost model to predict the likelihood of a customer responding to an email based on age, income, and location.

Why does the XGBoost model NOT meet the current requirements, and how can this be fixed?

Options:

A- The XGBoost model provides a true/false binary output. Apply principal component analysis (PCA) with five feature dimensions to predict a segment.

B- The XGBoost model provides a true/false binary output. Increase the number of classes the XGBoost model predicts to five classes to predict a segment.

C- The XGBoost model is a supervised machine learning algorithm. Train a k-Nearest-Neighbors (kNN) model with K = 5 on the same dataset to predict a segment.

D- The XGBoost model is a supervised machine learning algorithm. Train a k-means model with K = 5 on the same dataset to predict a segment.

Answer:

C

Question 9

Question Type: MultipleChoice

A retail company uses a machine learning (ML) model for daily sales forecasting. The company's brand manager reports that the model has provided inaccurate results for the past 3 weeks.

At the end of each day, an AWS Glue job consolidates the input data that is used for the forecasting with the actual daily sales data and the predictions of the model. The AWS Glue job stores the data in Amazon S3. The company's ML team is using an Amazon SageMaker Studio notebook to gain an understanding about the source of the model's inaccuracies.

What should the ML team do on the SageMaker Studio notebook to visualize the model's degradation MOST accurately?

Options:

A- Create a histogram of the daily sales over the last 3 weeks. In addition, create a histogram of the daily sales from before that period.

B- Create a histogram of the model errors over the last 3 weeks. In addition, create a histogram of the model errors from before that period.

C- Create a line chart with the weekly mean absolute error (MAE) of the model.

D- Create a scatter plot of daily sales versus model error for the last 3 weeks. In addition, create a scatter plot of daily sales versus model error from before that period.

Answer:

C

Question 10

Question Type: **MultipleChoice**

A data scientist has a dataset of machine part images stored in Amazon Elastic File System (Amazon EFS). The data scientist needs to use Amazon SageMaker to create and train an image classification machine learning model based on this dataset. Because of budget and time constraints, management wants the data scientist to create and train a model with the least number of steps and integration work required.

How should the data scientist meet these requirements?

Options:

A- Mount the EFS file system to a SageMaker notebook and run a script that copies the data to an Amazon FSx for Lustre file system. Run the SageMaker training job with the FSx for Lustre file system as the data source.

B- Launch a transient Amazon EMR cluster. Configure steps to mount the EFS file system and copy the data to an Amazon S3 bucket by using S3DistCp. Run the SageMaker training job with Amazon S3 as the data source.

C- Mount the EFS file system to an Amazon EC2 instance and use the AWS CLI to copy the data to an Amazon S3 bucket. Run the SageMaker training job with Amazon S3 as the data source.

D- Run a SageMaker training job with an EFS file system as the data source.

Answer:

A

Question 11

Question Type: MultipleChoice

A machine learning (ML) specialist wants to create a data preparation job that uses a PySpark script with complex window aggregation operations to create data for training and testing. The ML specialist needs to evaluate the impact of the number of features and the sample count on model performance.

Which approach should the ML specialist use to determine the ideal data transformations for the model?

Options:

A- Add an Amazon SageMaker Debugger hook to the script to capture key metrics. Run the script as an AWS Glue job.

B- Add an Amazon SageMaker Experiments tracker to the script to capture key metrics. Run the script as an AWS Glue job.

C- Add an Amazon SageMaker Debugger hook to the script to capture key parameters. Run the script as a SageMaker processing job.

D- Add an Amazon SageMaker Experiments tracker to the script to capture key parameters. Run the script as a SageMaker processing job.

Answer:

B