MLS-C01 Q&As

AWS Certified Machine Learning - Specialty (MLS-C01)

Pass Amazon MLS-C01 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.passapply.com/aws-certified-machine-learning-specialty.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Amazon Official Exam Center

QUESTION 1

A company has set up and deployed its machine learning (ML) model into production with an endpoint using Amazon SageMaker hosting services. The ML team has configured automatic scaling for its SageMaker instances to support workload changes. During testing, the team notices that additional instances are being launched before the new instances are ready. This behavior needs to change as soon as possible.

How can the ML team solve this issue?

A. Decrease the cooldown period for the scale-in activity. Increase the configured maximum capacity of instances.

B. Replace the current endpoint with a multi-model endpoint using SageMaker.

C. Set up Amazon API Gateway and AWS Lambda to trigger the SageMaker inference endpoint.

D. Increase the cooldown period for the scale-out activity.

Correct Answer: D

Reference: https://aws.amazon.com/blogs/machine-learning/configuring-autoscaling-inference-endpoints-in-amazon-sagemaker/

QUESTION 2

A wildlife research company has a set of images of lions and cheetahs. The company created a dataset of the images. The company labeled each image with a binary label that indicates whether an image contains a lion or cheetah. The company wants to train a model to identify whether new images contain a lion or cheetah.

Which Amazon SageMaker algorithm will meet this requirement?

A. XGBoost

B. Image Classification - TensorFlow

C. Object Detection - TensorFlow

D. Semantic segmentation - MXNet

Correct Answer: B

QUESTION 3

A machine learning (ML) specialist has prepared and used a custom container image with Amazon SageMaker to train an image classification model. The ML specialist is performing hyperparameter optimization (HPO) with this custom container image to produce a higher quality image classifier.

The ML specialist needs to determine whether HPO with the SageMaker built-in image classification algorithm will produce a better model than the model produced by HPO with the custom container image. All ML experiments and HPO jobs must be invoked from scripts inside SageMaker Studio notebooks.

How can the ML specialist meet these requirements in the LEAST amount of time?

A. Prepare a custom HPO script that runs multiple training jobs in SageMaker Studio in local mode to tune the model of the custom container image. Use the automatic model tuning capability of SageMaker with early stopping enabled to tune the model of the built-in image classification algorithm. Select the model with the best objective metric value.

B. Use SageMaker Autopilot to tune the model of the custom container image. Use the automatic model tuning capability of SageMaker with early stopping enabled to tune the model of the built-in image classification algorithm. Compare the objective metric values of the resulting models of the SageMaker AutopilotAutoML job and the automatic model tuning job. Select the model with the best objective metric value.

C. Use SageMaker Experiments to run and manage multiple training jobs and tune the model of the custom container image. Use the automatic model tuning capability of SageMaker to tune the model of the built-in image classification algorithm. Select the model with the best objective metric value.

D. Use the automatic model tuning capability of SageMaker to tune the models of the custom container image and the built-in image classification algorithm at the same time. Select the model with the best objective metric value.

Correct Answer: B

QUESTION 4

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.

Correct Answer: AC

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

QUESTION 5

A company is building a new supervised classification model in an AWS environment. The company\\'s data science team notices that the dataset has a large quantity of variables Ail the variables are numeric. The model accuracy for training and validation is low. The model\\'s processing time is affected by high latency The data science team needs to increase the accuracy of the model and decrease the processing.

How it should the data science team do to meet these requirements?

A. Create new features and interaction variables.

B. Use a principal component analysis (PCA) model.

C. Apply normalization on the feature set.

D. Use a multiple correspondence analysis (MCA) model

Correct Answer: B

The best way to meet the requirements is to use a principal component analysis (PCA) model, which is a technique that reduces the dimensionality of the dataset by transforming the original variables into a smaller set of new variables, called

principal components, that capture most of the variance and information in the data1. This technique has the following advantages:

It can increase the accuracy of the model by removing noise, redundancy, and multicollinearity from the data, and by enhancing the interpretability and generalization of the model23. It can decrease the processing time of the model by

reducing the number of features and the computational complexity of the model, and by improving the convergence and stability of the model45. It is suitable for numeric variables, as it relies on the covariance or correlation matrix of the data,

and it can handle a large quantity of variables, as it can extract the most relevant ones16. The other options are not effective or appropriate, because they have the following drawbacks:

A: Creating new features and interaction variables can increase the accuracy of the model by capturing more complex and nonlinear relationships in the data, but it can also increase the processing time of the model by adding more features

and increasing the computational complexity of the model7. Moreover, it can introduce more noise, redundancy, and multicollinearity in the data, which can degrade the performance and interpretability of the model8.

C: Applying normalization on the feature set can increase the accuracy of the model by scaling the features to a common range and avoiding the dominance of some features over others, but it can also decrease the processing time of the

model by reducing the numerical instability and improving the convergence of the model . However, normalization alone is not enough to address the high dimensionality and high latency issues of the dataset, as it does not reduce the

number of features or the variance in the data.

D: Using a multiple correspondence analysis (MCA) model is not suitable for numeric variables, as it is a technique that reduces the dimensionality of the dataset by transforming the original categorical variables into a smaller set of new

variables, called factors, that capture most of the inertia and information in the data. MCA is similar to PCA, but it is designed for nominal or ordinal variables, not for continuous or interval variables.

References:

1: Principal Component Analysis - Amazon SageMaker

2: How to Use PCA for Data Visualization and Improved Performance in Machine Learning | by Pratik Shukla | Towards Data Science

3: Principal Component Analysis (PCA) for Feature Selection and some of its Pitfalls | by Nagesh Singh Chauhan | Towards Data Science

4: How to Reduce Dimensionality with PCA and Train a Support Vector Machine in Python | by James Briggs | Towards Data Science

5: Dimensionality Reduction and Its Applications | by Aniruddha Bhandari | Towards Data Science

6: Principal Component Analysis (PCA) in Python | by Susan Li | Towards Data Science

7: Feature Engineering for Machine Learning | by Dipanjan (DJ) Sarkar | Towards Data Science

8: Feature Engineering -- How to Engineer Features and How to Get Good at It | by Parul Pandey | Towards Data Science

: [Feature Scaling for Machine Learning: Understanding the Difference Between Normalization vs. Standardization | by Benjamin Obi Tayo Ph.D. | Towards Data Science] : [Why, How and When to Scale your Features | by George Seif |

Towards Data Science] : [Normalization vs Dimensionality Reduction | by Saurabh Annadate | Towards Data Science] : [Multiple Correspondence Analysis - Amazon SageMaker] : [Multiple Correspondence Analysis (MCA) | by Raul Eulogio |

Towards Data Science]