

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.pass4itsure.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

Instant Download After Purchase

100% Money Back Guarantee

- 😳 365 Days Free Update
- 800,000+ Satisfied Customers





QUESTION 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? (Select TWO.)

A. Customize the built-in image classification algorithm to use Inception and use this for model training.

B. Create a support case with the SageMaker team to change the default image classification algorithm to Inception.

C. Bundle a Docker container with TensorFlow Estimator loaded with an Inception network and use this for model training.

D. Use custom code in Amazon SageMaker with TensorFlow Estimator to load the model with an Inception network and use this for model training.

E. Download and apt-get install the inception network code into an Amazon EC2 instance and use this instance as a Jupyter notebook in Amazon SageMaker.

Correct Answer: CD

QUESTION 2

A machine learning (ML) specialist needs to extract embedding vectors from a text series. The goal is to provide a readyto-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.)

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)

Correct Answer: AC

Reference: https://aws.amazon.com/blogs/machine-learning/create-a-word-pronunciation-sequence-to-sequence-model-using-amazon-sagemaker/ https://docs.aws.amazon.com/sagemaker/latest/dg/object2vec.html

QUESTION 3



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

Which prior probability distribution should the ML Specialist use for this variable?

- A. Poisson distribution
- B. Uniform distribution
- C. Normal distribution
- D. Binomial distribution
- Correct Answer: A

https://www.investopedia.com/terms/d/discrete-distribution.asp The Poisson distribution is commonly used for count data, which is the case here as we are interested in the number of minutes New Yorkers wait for a bus. The Poisson distribution is characterized by a single parameter, lambda, which represents the mean and variance of the distribution. In this case, the mean is 3 minutes, so we would set lambda to 3. The Poisson distribution assumes that events occur independently of each other, which is a reasonable assumption in this case since the waiting time for each individual is likely to be independent of the waiting time for others.

QUESTION 4

A machine learning engineer is building a bird classification model. The engineer randomly separates a dataset into a training dataset and a validation dataset. During the training phase, the model achieves very high accuracy. However, the model did not generalize well during validation of the validation dataset. The engineer realizes that the original dataset was imbalanced.

What should the engineer do to improve the validation accuracy of the model?

- A. Perform stratified sampling on the original dataset.
- B. Acquire additional data about the majority classes in the original dataset.
- C. Use a smaller, randomly sampled version of the training dataset.
- D. Perform systematic sampling on the original dataset.
- Correct Answer: A

Balanced Class Representation. Stratified sampling divides the original dataset into strata (groups) based on the class labels. It then selects instances from each stratum in a proportional manner, ensuring that the class distribution in the

training and validation datasets reflects the original class distribution. Improved Generalization. By having a balanced representation of all classes in the training and validation datasets, the model is exposed to a diverse range of instances

during training. This helps the model learn the distinguishing features of each class more effectively, leading to better generalization performance on the validation dataset.

Addressing Imbalanced Data. Stratified sampling directly addresses the issue of imbalanced data, which was identified as the root cause of the model/\'s poor generalization performance on the validation dataset.



QUESTION 5

A Machine Learning Specialist is preparing data for training on Amazon SageMaker The Specialist is transformed into a numpy .array, which appears to be negatively affecting the speed of the training

What should the Specialist do to optimize the data for training on SageMaker\\'?

- A. Use the SageMaker batch transform feature to transform the training data into a DataFrame
- B. Use AWS Glue to compress the data into the Apache Parquet format
- C. Transform the dataset into the Recordio protobuf format
- D. Use the SageMaker hyperparameter optimization feature to automatically optimize the data

Correct Answer: C

Most Amazon SageMaker algorithms work best when you use the optimized protobul recordIO format for the training data. https://docs.aws.amazon.com/sagemaker/latest/dg/cdf-training.html

MLS-C01 Study Guide

MLS-C01 Exam Questions

MLS-C01 Braindumps