Amazon

MLS-C01

AWS Certified Machine Learning - Specialty

QUESTION & ANSWERS

QUESTION 1

A Machine Learning Specialist observes several performance problems with the training portion of a machine learning solution on Amazon SageMaker The solution uses a large training dataset 2 TB in size and is using the SageMaker k-means algorithm The observed issues include the unacceptable length of time it takes before the training job launches and poor I/O throughput while training the model What should the Specialist do to address the performance issues with the current solution?

A. Use the SageMaker batch transform feature
B. Compress the training data into Apache Parquet format.
C. Ensure that the input mode for the training job is set to Pipe.
D. Copy the training dataset to an Amazon EFS volume mounted on the SageMaker instance.

Correct Answer: B

QUESTION 2

A Machine Learning Specialist is building a logistic regression model that will predict whether or not a person will order a pizza. The Specialist is trying to build the optimal model with an ideal classification threshold. What model evaluation technique should the Specialist use to understand how different classification thresholds will impact the model's performance?

A. Receiver operating characteristic (ROC) curve
B. Misclassification rate
C. Root Mean Square Error (RM&)
D. L1 norm

Correct Answer: A

QUESTION 3

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%r 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 4

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 long 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 5

A large consumer goods manufacturer has the following products on sale
• 34 different toothpaste variants
• 48 different toothbrush variants
• 43 different mouthwash variants
The entire sales history of all these products is available in Amazon S3 Currently, the company is using
custom-built autoregressive integrated moving average (ARIMA) models to forecast demand for these products The company wants to predict the demand for a new product that will soon be launched Which solution should a Machine Learning Specialist apply?

A. Train a custom ARIMA model to forecast demand for the new product.

B. Train an Amazon SageMaker DeepAR algorithm to forecast demand for the new product

C. Train an Amazon SageMaker k-means clustering algorithm to forecast demand for the new product.

D. Train a custom XGBoost model to forecast demand for the new product

Explanation/Reference:

The Amazon SageMaker DeepAR forecasting algorithm is a supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNN). Classical forecasting methods, such as autoregressive integrated moving average (ARIMA) or exponential smoothing (ETS), fit a single model to each individual time series. They then use that model to extrapolate the time series into the future.

QUESTION 6

A Machine Learning Specialist has built a model using Amazon SageMaker built-in algorithms and is not getting expected accurate results The Specialist wants to use hyperparameter optimization to increase the model's accuracy Which method is the MOST repeatable and requires the LEAST amount of effort to achieve this?

A. Launch multiple training jobs in parallel with different hyperparameters
B. Create an AWS Step Functions workflow that monitors the accuracy in Amazon CloudWatch Logs and relaunches the training job with a defined list of hyperparameters
C. Create a hyperparameter tuning job and set the accuracy as an objective metric.
D. Create a random walk in the parameter space to iterate through a range of values that should be used for each individual hyperparameter

QUESTION 7

A Data Science team within a large company uses Amazon SageMaker notebooks to access data stored in Amazon S3 buckets. The IT Security team is concerned that internet-enabled notebook instances create a security vulnerability where malicious code running on the instances could compromise data privacy. The company mandates that all instances stay within a secured VPC with no internet access, and data communication traffic must stay within the AWS network. How should the Data Science team configure the notebook instance placement to meet these requirements?

A. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Place the Amazon SageMaker endpoint and S3 buckets within the same VPC.
B. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Use 1AM policies to grant access to Amazon S3 and Amazon SageMaker.
C. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has S3 VPC endpoints and Amazon SageMaker VPC endpoints attached to it.
D. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has a NAT gateway and an associated security group allowing only outbound connections to Amazon S3 and Amazon SageMaker

QUESTION 8

While reviewing the histogram for residuals on regression evaluation data a Machine Learning Specialist notices that the residuals do not form a zero-centered bell shape as shown What does this mean?
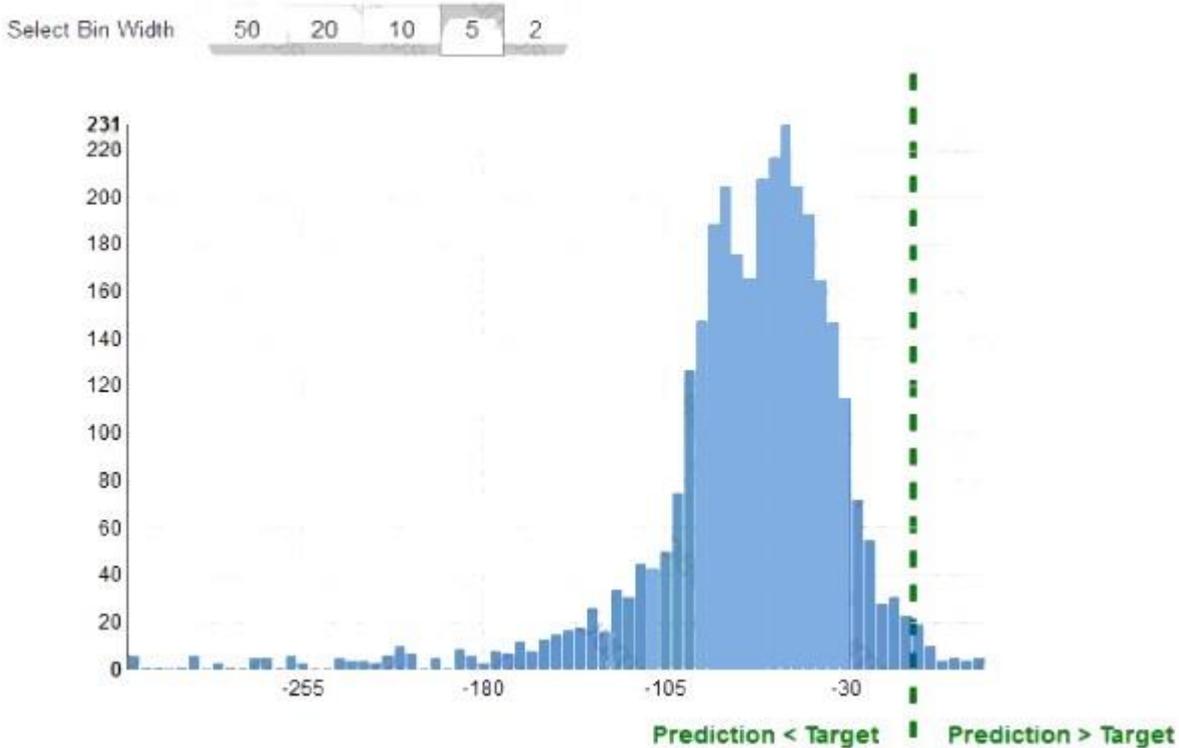
A. The model might have prediction errors over a range of target values.
B. The dataset cannot be accurately represented using the regression model
C. There are too many variables in the model
D. The model is predicting its target values perfectly.

QUESTION 9

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

n = 100	PREDICTED CHURN Yes	PREDICTED CHURN No
ACTUAL Churn Yes	10	4
Actual No	10	76

A. The precision of the model is 86%, which is less than the accuracy of the model.
B. The precision of the model is 86%, which is greater than the accuracy of the model.

Correct Answer:B

QUESTION 10

A Data Scientist is developing a machine learning model to classify whether a financial transaction is fraudulent. The labeled data available for training consists of 100,000 non-fraudulent observations and 1,000 fraudulent observations. The Data Scientist applies the XGBoost algorithm to the data, resulting in the following confusion matrix when the trained model is applied to a previously unseen validation dataset. The accuracy of the model is 99.1%, but the Data Scientist has been asked to reduce the number of false negatives.

Predicted	0	1
Actual 0	99,966	34
1	877	123

Which combination of steps should the Data Scientist take to reduce the number of false positive predictions by the model? (Select TWO.)

A. Change the XGBoost eval_metric parameter to optimize based on rmse instead of error.
B. Increase the XGBoost scale_pos_weight parameter to adjust the balance of positive and negative weights.
C. Increase the XGBoost max_depth parameter because the model is currently underfitting the data.
D. Change the XGBoost evaljnetric parameter to optimize based on AUC instead of error.
E. Decrease the XGBoost max_depth parameter because the model is currently overfitting the data.

Correct Answer:D,E