

Special Offer | Flat 15% OFF on All Courses | Use Coupon - WHIZSITE15

[Home](#) > [My Courses](#) > [AWS Certified Machine Learning Specialty](#) > [Modeling](#) > [Report](#)

Search Courses



## Modeling

Completed on 28-January-2021



Attempt

01



Marks Obtained

0 / 10



Your score

0.0%



Time Taken

00 H 00 M 49 S



Result

Failed

### Domains wise Quiz Performance Report

Join us on [Slack community](#)

No	Domain	Total Question	Correct	Incorrect	Unattempted	Marked as Review
1	<a href="#">Modeling</a>	10	0	1	9	0
Total	<a href="#">All Domain</a>	10	0	1	9	0

[Review the Answers](#)

Sorting by

All

### Question 1

Unattempted

Domain :Modeling

You are a machine learning specialist at a large financial services firm. You are building a machine learning model to manage risk for your firm using data from your traders daily trading activity. You are in the stage of your model development where you need to provide jupyter notebooks to your development team that they can use in SageMaker Studio. Your developers need to use the Scala kernel based on the Almond Scala Kernel as their development environment in their jupyter notebooks.

How can you provide the required development environment to your developers for their jupyter notebooks in the most efficient manner?

- A. Allow the developers to select the Scala kernel based on the Almond Scala Kernel from the list of included SageMaker images available in SageMaker studio

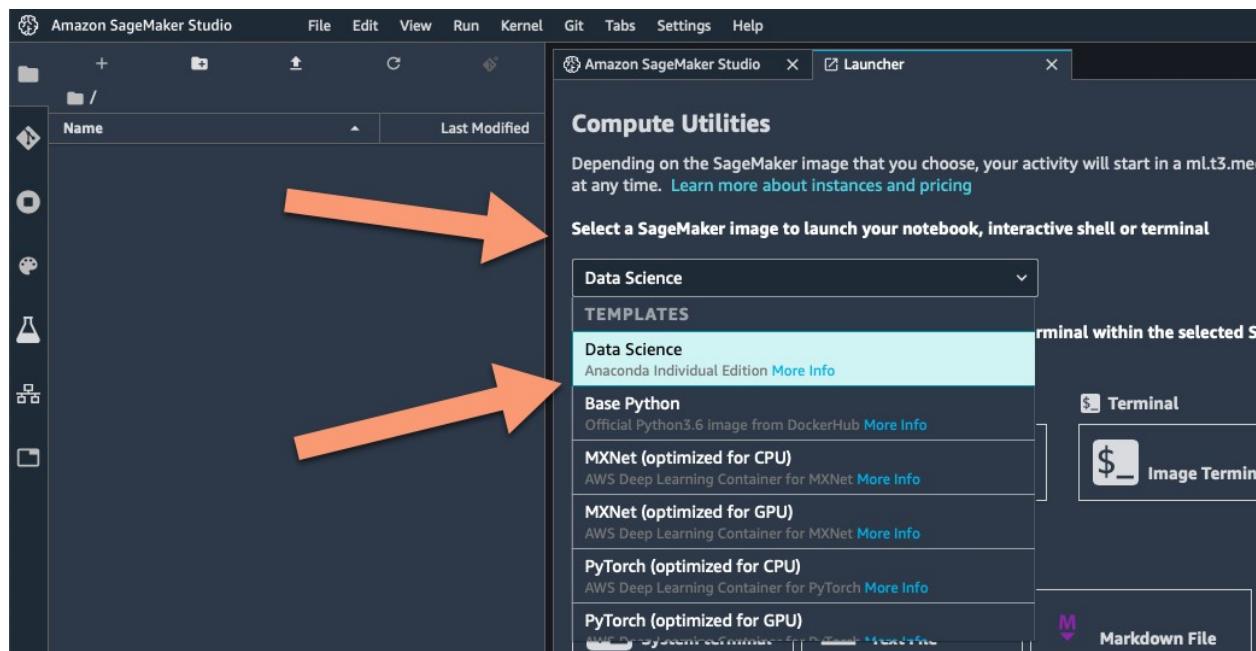
- B. Create a custom SageMaker image based on an AMI that includes the Scala kernel based on the Almond Scala Kernel and attach the image to your SageMaker domain
- C. Create a custom SageMaker docker image using the Scala kernel based on the Almond Scala Kernel SageMaker custom image sample and attach the image to each user's profile
- D. Create a custom SageMaker docker image using the Scala kernel based on the Almond Scala Kernel SageMaker custom image sample and attach the image to the SageMaker domain



### Explanation:

Answer: D

**Option A is incorrect.** The default list of available images in SageMaker studio does not include a scala image.



**Option B is incorrect.** SageMaker Studio images are built on Docker images, not AMIs.

**Option C is incorrect.** While you should create a custom SageMaker docker image using the Scala kernel based on the Almond Scala Kernel SageMaker custom image sample, attaching it to each user's profile is less efficient than attaching it to the SageMaker Studio domain.

**Option D is CORRECT.** The most efficient option is to create a custom SageMaker docker image using the Scala kernel based on the Almond Scala Kernel SageMaker custom image sample and attach it to the SageMaker Studio domain.

### Reference:

Please see the **Amazon Sagemaker developer guide** titled [Bring your own SageMaker image](#).

Please refer to the GitHub repository titled [SageMaker Studio Custom Image Samples](#).

Ask our Experts

Rate this Question?  

View Queries

open ▾

### Question 2

Unattempted

Domain :Modeling

You work as a machine learning specialist for a city that wants to monitor its air quality to address air pollution in their environment. You and your machine learning specialist team need to forecast the city air quality in parts per million of contaminants over the next week, taking into account weather, traffic conditions, and other pollutant contributors. You are building your model using daily data from the last year as your data source. Your team has decided to use SageMaker Studio to leverage its collaborative notebooks feature.

Which model and SageMaker Studio image will provide the best results for your team in the most efficient manner?

- A. Use the SageMaker Studio Base TensorFlow [tensorflow-2.3.0] image and the k-Nearest-Neighbors algorithm on the single time series consisting of the full year of data with a predictor\_type of . regressor
- B. Use SageMaker Studio Base R [r-4.0.3] image and the Random Cut Forest algorithm on the single time series consisting of the full year of data.
- C. Use the SageMaker Studio Base Python [python-3.6] image and the Linear Learner algorithm on the single time series consisting of the full year of data with a predictor\_type of regressor. 
- D. Use the SageMaker Studio Base Scala [scala-2.13.3] image and the Linear Learner algorithm on the single time series consisting of the full year of data with a predictor\_type of classifier.

### Explanation:

Answer: C

**Option A is incorrect.** The problem of forecasting the air quality is a regression problem best solved by the Linear Learner built-in algorithm, not the k-Nearest Neighbors SageMaker built-in algorithm. Also, there is no Base TensorFlow image available in SageMaker Studio. You would have to create a custom image to use TensorFlow in

SageMaker Studio. This would be far less efficient than using an image that is already available with SageMaker Studio.

**Option B is incorrect.** The problem of forecasting the air quality is a regression problem best solved by the Linear Learner built-in algorithm, not the Random Cut Forest SageMaker built-in algorithm. Also, there is no Base R image available in SageMaker Studio. You would have to create a custom image to use R in SageMaker Studio. This would be far less efficient than using an image that is already available with SageMaker Studio.

**Option C is CORRECT.** The Linear Learner SageMaker built-in algorithm is the best choice from the options given to solve this regression problem. Also, the SageMaker Studio Base Python [python-3.6] image is a valid choice of an available SageMaker Studio image.

**Option D is incorrect.** While the Linear Learner SageMaker built-in algorithm is the best choice from the options given to solve this regression problem, there is no Base Scala image available in SageMaker Studio. You would have to create a custom image to use Scala in SageMaker Studio. This would be far less efficient than using an image that is already available with SageMaker Studio.

#### Reference:

Please see the [AWS SageMaker developer guide](#) titled [Available Amazon SageMaker Images](#).

Please refer to the [AWS SageMaker developer guide](#) titled [Bring your own SageMaker image](#).

---

Ask our Experts

Rate this Question?  

[View Queries](#)

open ▾

#### Question 3

Unattempted

Domain :Modeling

You are a machine learning specialist working for a large retail clothing company. Your marketing department wants to leverage machine learning to understand product loyalty. Your machine learning team has decided to group the customers into categories based on which customers may churn, meaning abandon a given product for another, maybe a competitor's product, within the next 6 months. You have labeled data for customer product loyalty for the previous two years available to you for training your model.

Which machine learning model type should you and your team use to build your customer loyalty model?

- A. Linear regression using either the Linear Learner or XGBoost built-in SageMaker algorithms
- B. Clustering using either the K-Means or K-Nearest Neighbors algorithms
- C. Classification using either the Linear Learner or XGBoost built-in SageMaker algorithms 
- D. Reinforcement learning using SageMaker RL

#### Explanation:

Answer: C

**Option A is incorrect.** Because you have labeled data and you want to group your customers into categories, classification is a better choice than regression.

**Option B is incorrect.** Clustering is done using unlabeled data. You have labeled data so you will get better results using a classification algorithm.

**Option C is CORRECT.** Since your data is labeled you will get the best results using a classification algorithm when attempting to classify your customers.

**Option D is incorrect.** Reinforcement algorithms are used for solving problems such as supply chain management, HVAC systems, industrial robotics, game artificial intelligence, dialog systems, and autonomous vehicles. They attempt to learn a strategy. They are not well suited to grouping data elements into categories.

#### Reference:

Please see the [Amazon SageMaker developer guide](#) titled [Use Amazon SageMaker Built-in Algorithms](#).

Please refer to the [article](#) titled [Regression vs Classification in Machine Learning: What is The Difference?](#).

Please review the [article](#) titled [How to Use Unlabeled Data in Machine Learning](#).

Please refer to the [Amazon SageMaker developer guide](#) titled [Use reinforcement learning with Amazon SageMaker](#).

Ask our Experts

Rate this Question?  

View Queries

open 

**Question 4****Unattempted****Domain :Modeling**

Your machine learning team is building and planning to operationalize a machine learning model that uses deep learning to recognize and classify images of potential security threats to important government officials using stills taken from live security video. However, when your team runs their mini-batch training of the neural network, the training accuracy oscillates over your training epochs.

What is the most probable cause of your training accuracy problem?

- A. The validation error has stopped decreasing
- B. The mini-batch size is too small
- C. The mini-batch size is too large
- D. The learning rate is very high 

**Explanation:**

Answer: D

**Option A is incorrect.** In deep learning training, choosing the correct number of epochs is important. The validation error is used to determine how many epochs to run through. When the learning rate stops decreasing you should stop running training epochs. The point when your validation error stops decreasing has no correlation to the oscillation of the accuracy of your training epochs.

**Option B is incorrect.** A small mini-batch is used to prevent your training process from stopping at local minima. Having a small mini-batch size won't cause oscillation in your training epoch accuracy.

**Option C is incorrect.** A large mini-batch size is used to allow for highly computational demanding matrix multiplication in your training calculations. Having a large mini-batch size won't cause oscillation in your training epoch accuracy.

**Option D is CORRECT.** A learning rate that is very high tends to cause oscillation in your training accuracy. A high learning rate causes your weight updates to be too large and you will overestimate your goal and oscillate around the true goal.

**Reference:**

Please see the [Amazon SageMaker developer guide](#) titled [DeepAR Forecasting Algorithm](#).

Please refer to the [Machine Learning Mastery article](#) titled [How to Configure the Learning Rate When Training Deep Learning Neural Networks](#).

Please review the **article** titled [Hyperparameters in Machine /Deep Learning](#).

Please refer to the **Towards Data Science article** titled [Hyper-parameter Tuning Techniques in Deep Learning](#).

---

Ask our Experts

Rate this Question?  

**View Queries**

open ▾

**Question 5**

**Unattempted**

Domain :Modeling

You are a machine learning specialist on a software development team in a real estate company. Your management team has asked your team to build a logistic regression model that your company wishes to use to predict whether or not a person will buy a given listing based on multiple attributes of the sale, the property, and the customer profile. Your team lead has assigned your team the task of finding the optimal model with an ideal classification threshold.

Which model evaluation technique should your team use to discover how different classification thresholds will affect the model's performance?

- A. Rand Index
- B. Root Mean Square Error (RMSE)
- C. Mean Absolute Error (MAE)
- D. Receiver Operating Characteristic (ROC) curve 

---

**Explanation:**

Answer: D

**Option A is incorrect.** The Rand Index evaluation technique is used for optimizing unsupervised models, not supervised models like logistic regression models.

**Option B is incorrect.** The RMSE evaluation technique is used for regression problems where you are solving for a continuous variable. In this use case you are solving for a binary classification: will or will not buy.

**Option C is incorrect.** The MAE evaluation technique is also used for regression problems where you are solving for a continuous variable. In this use case you are solving for a binary classification: will or will not buy.

**Option D is CORRECT.** The Receiver Operating Characteristic curve evaluation technique is used for regression problems where you are solving for a binary variable. In this use case you are solving for a binary classification: will or will not buy.

#### Reference:

Please see the [Google Machine Learning Crash Course article](#) titled [Classification: ROC Curve and AUC](#).

Please refer to the [Towards Data Science article](#) titled [Understanding AUC - ROC Curve](#).

Please review the [Data Institute article](#) titled [Choosing the Right Metric for Evaluating Machine Learning Models — Part 1](#).

[Ask our Experts](#)

Rate this Question?  

[View Queries](#)

open ▾

#### Question 6

Unattempted

Domain :Modeling

You are a machine learning specialist at a pharmaceutical drug manufacturer. Your team has the task of building a deep learning model to be used for drug discovery by combining data from a variety of sources. Your team will use the deep learning model to predict novel candidate biomolecules for new disease targets such as COVID 19. You and your team have created a deep learning neural network disease target prediction model that performs well on the training data, however, it performs poorly on your test data.

Which of the methods listed should your team use to correct your testing problem? (SELECT THREE)

- A. Increase dropout 
- B. Decrease regularization
- C. Increase regularization 
- D. Decrease dropout
- E. Decrease feature combinations 
- F. Increase feature combinations

#### Explanation:

**Answers: A, C and E**

**Option A is CORRECT.** When your model performs well in training but poorly in testing, it quite often means the model is overfitted or not generalized enough. Increasing dropout, or the probability of a node to be turned off, helps generalization.

**Option B is incorrect.** You need to increase regularization to avoid overfitting, a common problem when your training data is "memorized" by your neural network. Decreasing regularization is used when your model is underfitting.

**Option C is CORRECT.** You need to increase regularization to avoid overfitting, a common problem when your training data is "memorized" by your neural network. Increasing regularization helps generalization.

**Option D is incorrect.** Decreasing dropout will not help generalization, it will have the opposite effect.

**Option E is CORRECT.** Decreasing feature combinations, combining the strength of multiple complementary features to yield more powerful features (reducing the number of features), will help with generalization.

**Option F is incorrect.** Increasing feature combinations is used when your model is underfitting. Decreasing feature combinations, combining the strength of multiple complementary features to yield more powerful features (reducing the number of features), will help with generalization.

**Reference:**

Please see the [Amazon Machine Learning developer guide](#) titled **Model Fit: Underfitting vs. Overfitting**.

Please refer to the [Machine Learning Mastery article](#) titled **A Gentle Introduction to Dropout for Regularizing Deep Neural Networks**

[Ask our Experts](#)[Rate this Question?](#)  [View Queries](#)[open ▾](#)**Question 7****Unattempted****Domain :Modeling**

You work as a machine learning specialist for a publishing company. The company has labeled a historical dataset of publication sales data. Using this labeled data, you need to predict how many copies of a given publication should be produced each month.

Which machine learning algorithm type should you use to generate your predictions?

- A. Linear regression 
- B. Principal component analysis
- C. Random Cut Forest
- D. Logistic regression

---

**Explanation:**

Answer: A

**Option A is CORRECT.** You are trying to solve for a "how many" question and your data is labeled. These two factors lead to the choice of linear regression as the best option from those given.

**Option B is incorrect.** Principal component analysis is used for dimensionality reduction, not for solving predictions of "how many" problems. Also, it is an unsupervised algorithm. We have labeled data so we should use a supervised algorithm.

**Option C is incorrect.** Random cut forest is used primarily as an unsupervised algorithm for detecting anomalous data points within a data set. Since we have labeled data we will use a supervised algorithm.

**Option D is incorrect.** Logistic regression is used to solve "yes/no" or binary predictions, not "how many" predictions.

**Reference:**

Please see the [Amazon Machine Learning developer guide](#) titled [Regression Model Insights](#).

Please refer to the [Amazon SageMaker developer guide](#) titled [Random Cut Forest \(RCF\) Algorithm](#).

Please refer to the [Amazon SageMaker developer guide](#) titled [Principal Component Analysis \(PCA\) Algorithm](#)

---

Ask our Experts

Rate this Question?  

[View Queries](#)

open 

**Question 8**

**Unattempted**

Domain :Modeling

You are the technical leader of your machine learning team where you are responsible for the quality of your team's code. Your team generates a very high level of code per week, thousands of lines of Java code used in machine learning data ingestion, transformation, as well as EMR cluster jobs. You need a way to improve the pull request review process so you have decided to use CodeGuru Reviewer to leverage its program analysis and machine learning to detect defects that your developers may fail to find while also leveraging its suggestions for improving the team's Java code. To this end, you want to use CodeGuru to create code reviews. You are using GitHub Enterprise Server for your code repository. When your developers commit code to your GitHub repository you want the CodeGuru Reviewer code review to run. What do you need to do to make this workflow possible?

- A. Perform your own input validation for your developers code
- B. Select the type of regression SageMaker built-in algorithm to use in the machine learning defect detection
- C. Select the type of clustering SageMaker built-in algorithm to use in the machine learning defect detection
- D. Create a connection to your repository using AWS CodeStar connections to connect the CodeGuru Reviewer service to your GitHub repository



**Explanation:**

Answer: D

**Option A is incorrect.** CodeGuru Reviewer identifies problems with input validation for you using its machine learning capabilities..

**Option B is incorrect.** You don't need to select the machine learning algorithm used by CodeGuru Reviewer.

**Option C is incorrect.** You don't need to select the machine learning algorithm used by CodeGuru Reviewer.

**Option D is CORRECT.** If you use GitHub Enterprise Server as your code repository, you are required to create a connection to your repository using AWS CodeStar connections to connect the CodeGuru Reviewer service to your GitHub repository.

**Reference:**

Please see the [Amazon CodeGuru Reviewer user guide](#) titled [What is Amazon CodeGuru Reviewer?](#).

Please see the [Amazon CodeGuru Reviewer user guide](#) titled [Create a repository for your source code](#).

Please see the [Developer Tools console user guide](#) titled [What are connections?](#).

---

Ask our Experts

Rate this Question?  

---

**View Queries**

open ▾

---

**Question 9**

**Unattempted**

Domain :Modeling

You are a machine learning specialist at a large bank. Your machine learning team has recently been assigned the task of detecting fraud in the bank's web and mobile applications. Your management team is excited about using machine learning for fraud detection but they have limited money in the yearly budget for this work.

You have decided to use the Amazon Fraud Detector service to deliver your fraud detection layer in your web and mobile architecture. When building your Fraud Detector model, which model type should you choose?

- A. **ONLINE\_FRAUD\_DETECTOR**

B. ONLINE\_FRAUD\_INSIGHTS 

C. FRAUD\_INSIGHTS

D. FRAUD\_DECTOR

---

**Explanation:**

Answer: B

**Option A is incorrect.** The model type available for the Fraud Detector service is the ONLINE\_FRAUD\_INSIGHTS model.

**Option B is CORRECT.** The model available in the Fraud Detector service is the ONLINE\_FRAUD\_INSIGHTS model.

**Option C is incorrect.** The model type available for the Fraud Detector service is the ONLINE\_FRAUD\_INSIGHTS model.

**Option D is incorrect.** The model type available for the Fraud Detector service is the ONLINE\_FRAUD\_INSIGHTS model.

**Reference:**

Please see the [Amazon Fraud Detector user guide](#) titled [How Amazon Fraud Detector works](#).

Please see the [Amazon Fraud Detector welcome page](#) titled [CreateModel](#).

Please see the [Amazon Fraud Detector FAQs](#).

---

Ask our Experts

Rate this Question?  

[View Queries](#)

open ▾

**Question 10**

**Incorrect**

Domain :Modeling

Your machine learning team is part of the research department of a hedge fund firm. Your team has been assigned a project to forecast the price movement of several stocks in the NASDAQ index. You have decided to use historical related time series in your model to improve the accuracy of your model. Your management team has asked that your team produce the model quickly and at a low administrative overhead so your team lead has decided to use the Amazon Forecast service

Which Amazon Forecast algorithm would be the best choice for your stock price movement forecasting problem?

- A. Prophet
- ✓ B. DeepAR+ 
- C. ARIMA
- D. CNN-QR 

---

**Explanation:**

Answer: D

**Option A is incorrect.** The Amazon Forecast Prophet algorithm does not accept related time series data without future values.

**Option B is incorrect.** The Amazon Forecast DeepAR+ algorithm does not accept related time series data without future values.

**Option C is incorrect.** The Amazon Forecast ARIMA algorithm does not accept related time series data without future values.

**Option D is CORRECT.** The Amazon Forecast CNN-QR algorithm is the only Forecast algorithm that accepts related time series data without future values.

**Reference:**

Please see the [Amazon Forecast developer guide](#) titled [Getting Started](#).

Please see the [Amazon Forecast developer guide](#) titled [Choosing an Amazon Forecast Algorithm](#).

Please see the [Amazon Forecast developer guide](#) titled [Using Related Time Series Datasets](#).

---

Ask our Experts

Rate this Question?  

---

**View Queries**

open 

Finish Review

Certification	Company	Support	Join us on Slack!
Cloud Certification	Become Our Instructor	Contact Us	 Join our open <b>Slack community</b> and get your queries answered instantly! Our experts are online to answer your questions!
Java Certification	Support	Help Topics	
PM Certification	Discussions		
Big Data Certification	Blog		<b>Follow us</b>
	Business		  

---

© Copyright 2021. Whizlabs Software Pvt. Ltd. All Right Reserved.