

Cloud Guru - 6 - Algorithms Quiz

<https://acloud.guru/course/aws-certified-machine-learning-specialty/learn/1efec92d-271f-f097-0c26-236b41ab42b1/chapter-7/watch?backUrl=%2Fcourses>

Let's test your knowledge!

AWS Certified Machine Learning - Specialty 2020 - Algorithms Quiz

ABOUT THIS QUIZ

NO. OF QUESTIONS
14 Questions

SKILL LEVEL
Intermediate

QUESTION 1

Which best describes SGD in common terms?

- ☐ Attempt to find the most efficient path to deliver packages to multiple destinations.
- ☒ Seek to find the lowest point in elevation on a landscape.
- ☐ Ensure that our sample size in a traffic study has at least 30 drivers.
- ☐ Calculate the linear distance between arrows shot into a target to determine accuracy.

Good work!

Stochastic Gradient Descent (SGD) is a cost function that seeks to find the minimal error. This can be analogous to trying to find the lowest point on a landscape.

QUESTION 2

To remove inconsistency in a process, you have created a very specific step-by-step process for patching servers. How would this best be described?

- ☒ The procedure is an algorithm as it will yield a consistent output.
- ☐ The procedure is an algorithm as it will allow the person performing the upgrades to choose the best path.
- ☐ The procedure is a work-around until a supervised learning process can be implemented which will remove the human from the process.
- ☐ The procedure is a heuristic as it will yield a consistent output.
- ☐ The procedure is a heuristic as it permits the person performing the upgrades from making unintended mistakes.

Good work!

An algorithm is a specific set of steps intended to solve a specific problem with consistency. A heuristic is mental "rule of thumb" that provides guidance but does not always ensure a consistent outcome.

QUESTION 3

While using K-Means, what does it mean if we pass in $k=4$ as a hyperparameter?

- ☐ We want the algorithm to classify into 4 groups.
- ☐ We want the algorithm to return the top 4 results.
- ☐ We want the algorithm to use 4 as the cutoff value for classification purposes.
- ☒ We want the algorithm to group into 4 clusters.
- ☐ We want the algorithm to group into clusters of no more than 4 samples each.

Good work!

K-Means is a clustering algorithm and the K represents the number of clusters the algorithm should seek.

QUESTION 4

You are on a personal quest to design the best chess playing model in the world. What might be a good strategy for this objective?

- ☐ Use an unsupervised learning strategy to analyze similarities across thousands of the best chess matches.
- ☐ Use Mechanical Turk to crowdsource the best chess moves in a variety of scenarios then use those for a supervised learning session.
- ☒ Use a reinforcement learning strategy to let the model learn itself.
- ☐ Use a Factorization Machine approach to analyze the winning series of moves across thousands of chess matches to find the best series of moves.
- ☐ Use a supervised learning strategy that is trained by feeding in the chess moves of thousands of famous chess experts.

Good work!

Chess is a complex game and would require some training to develop a good model. Supervised learning, by definition, can only be as good as the training data that it is supplied with so we won't be able to meet our goal of the best chess model using supervised learning. Rather, reinforcement learning would have the best chance of becoming better than any other player if given enough iterations and a proper reward function.

QUESTION 5

You have been asked to help develop a vision system for a manufacturing line that will reorient parts to a specific position using a robotic arm. What algorithm might you choose for the vision part of this problem?

☐ Seq2Seq

☐ AWS Comprehend

☐ Object2Vec

☒ Image Analysis

☐ Object Detection

☒ Semantic Segmentation

Sorry!

Correct Answer

Semantic Segmentation can perform edge detection, which could be used to identify orientation. Object Detection and Image Analysis are better used for full images where as Semantic Segmentation can create a segmentation mask or outline of a part of the image.

QUESTION 6

In what scenario is the DeepAR algorithm best suited?

- ☐ Decide whether to extend a credit card offer to a potential customer.
- ☐ Predict whether a football team will score a certain number of points in a match.
- ☒ Predict future sales of a new product based on historic sales of similar products.
- ☐ Provide the certainty that a given picture includes a human face.
- ☐ Determine the correlation between a person's diet and energy levels.

Good work!

DeepAR is a supervised forecasting algorithm used with time-series data. DeepAR seeks to be better than traditional time-series algorithms by accommodating multiple cross-sectional datasets.

QUESTION 7

When analyzing a set of one-hot encoded data you realize that, while there is a massive amount of data, most of the values are absent. This is expected given the type of data, but what built-in SageMaker algorithm might you choose to work with this data?

- ☐ Object2Vec
- ☒ Factorization Machines
- ☐ Linear Learner
- ☐ Semantic Segmentation
- ☐ K-Nearest Neighbor

Good work!

From the description, it seems that we have a very sparse dataset. Factorization Machines is most often associated with high-dimensional sparse datasets that we are likely to see in one-hot encoded datasets.

QUESTION 8

Which of the following might be used to focus a model on most relevant features?

☐ NLM

☐ AQS

☐ K-NN

☐ XGB

☐ LDA

☒ PCA

Good work!

Principal Component Analysis (PCA) is an unsupervised algorithm that attempts to reduce dimensionality while retaining as much information as possible.

QUESTION 9

You are consulting for a shipping company who wants to implement a very specific algorithm for shipping container optimization. The algorithm is not part of the currently available SageMaker built-in algorithms. What are your options?

Choose 2



Build the algorithm in a docker container and use that custom algorithm for training and inference in SageMaker.



Use a series of existing algorithms to simulate the actions of the unavailable algorithm.



Wait until the algorithm is available in SageMaker before further work.



Post an incendiary message to Twitter hoping to shame AWS into adopting the specialized algorithm.



Search the AWS Marketplace for the algorithm. If it exists, deploy it using SageMaker for inferences.

Good work!

If SageMaker does not support a desired algorithm, you can either bring your own or buy/subscribe to an algorithm from the AWS Marketplace.

QUESTION 10

You are being asked to develop a model to predict the likelihood that a student will pass a certification exam based on hours of study. Of the options given, what would be the best approach to this problem?

- ☐ Build a simulation-based model which will analyze past student performance at varying levels of study.
- ☐ Build a clustering mode with K-Means to group students who pass in a cluster.
- ☐ Build a model using LDA based on the text of the questions on the exam and predict student outcome.
- ☐ Build a model using NLP to classify students into passing and failing groups.
- ☒ Build a Logistic Regression model using the hours of study as as a feature.

Good work!

This problem is best described as a binary classification problem as we are trying to understand whether a student will pass or fail. The option that most directly provides for a binary classification problem is logistic regression.

QUESTION 11

Which of the following is an example of unsupervised learning?

Choose 2

- ☒ Using K-Means to cluster customers into demographic segments.
- ☒ Using NTM to extract topics from a set of scientific journal articles.
- ☐ Using a Factorization Machine to provide book recommendations.
- ☐ Using XGBoost to predict the selling price of a house in a particular market.
- ☐ Using Seq2Seq to extract a text string from a segment of a recorded speech.

Good work!

Both K-Means and Neural Topic Modelling are unsupervised learning methods. XGBoost, Seq2Seq, and Factorization Machines are supervised learning methods.

QUESTION 12

Which of these examples would be considered as introducing bias into a problem space?

Choose 2

- ☒ Failing to randomize a dataset even though you were told it was already random.
- ☐ Deciding to use a supervised learning method to estimate missing values in a dataset.
- ☒ Omitting records before a certain date in a forecasting problem.
- ☐ Filtering out outliers in a dataset which are greater than 4 standard deviations outside the mean.
- ☒ Removing records from a set of customer reviews that were not fully complete.

Sorry!

Correct Answer

Failing to randomize records even when you think they are already random can lead to introducing bias in the model when splitting. Additionally, removing customer review records could unintentionally filter out statistically significant parts of the dataset, biasing towards those customers who have the time or energy to fully complete the reviews. The other answers are reasonable and do not explicitly introduce bias.

QUESTION 13

Which of the following are good candidate problems for using XGBoost?

Choose 3



Evaluate handwritten numbers on a warranty card to detect what number they represent.



Map a text string to an n-gram vector.



Deciding whether a transaction is fraudulent or not based on various details about the transaction.



Create a policy that will guide an autonomous robot through an unknown maze.



Providing a ranking of search results on an e-commerce site customized to a customer's past purchases.

Good work!

XGBoost is an extremely flexible algorithm which can be used in regression, binary classification, and multi-class classification.

QUESTION 14

You are consulting with a retailer that wants to evaluate the sentiment of social media posts to determine if they are positive or negative. Which approach would be the most direct to this problem?

- ☐ Use BlazingText in Word2Vec mode for skip-gram.
- ☐ Use Amazon Macie.
- ☒ Use Amazon Comprehend.
- ☐ Use Object2Vec in sentiment detection mode.
- ☐ Use BlazingText in Text Classification mode.

Good work!

The most direct method of sentiment analysis is using the Amazon Comprehend service. Word2Vec can be used as a pre-processing step but it alone is not sufficient to detect sentiment.

86%

Congratulations!

You passed AWS Certified Machine Learning - Specialty 2020 - Algorithms Quiz!