SuperDataScience

10 SAMPLE EXAM QUESTIONS
+ COMPLETE ANSWER KEY

MACHINE AND DEEP LEARNING IN AWS — PART #2

1. **John is building an image classifier to differentiate Square from triangles. He has both the training images and their corresponding labels. Which Scikit-learn function would help him classify?**

   A SVM (Support Vector Machine)
   B K-means
   C PCA
   D Grid search

2. **You are working on sentiment analysis model for tweets. You want to analysis the sentiment for a client company. So, you want to collect the tweets containing the client company name and store in S3. How can you go about it?**

   A Use Kinesis Data Stream to add the data to S3 and use Lambda function to trigger for the particular criteria when the data is S3
   B Use Kinesis Data Stream to add the data to S3 and use Kinesis Analytics to trigger for the particular criteria when the data is S3
   C Use Kinesis Firehose to add the data to S3 and use Lambda function to trigger for the particular criteria when the data is S3
   D Use Kinesis Firehose to add the data to S3 and use Kinesis Analytics to trigger for the particular criteria when the data is S3

3. **John is given access to data from a real-estate company to create model to predict the price of the house in a particular area. This data contains more than 50 features and he was looking for way to reduce the dimensionality of the data, to get the most relevant features?**

   A XGBoost
   B MLP

C PCA

D LDA

4. **You want to cluster the customers of an e-commerce company based on the demographic region of their IP. Which algorithm would help with this issue?**

A Use AWS Comprehend

B Use K-means

C Use AWS Rekognition

D Use Factorization Machine

5. **You have given access to large set of data and you are tasked to group the features based on their similarity which of the following would help you to group the similar data?**

A K-means

B Random Forest

C PCA

D Support Vector Machines

E MLP

6. **Jack is working on a machine learning problem using Sagemaker built-in algorithm and he wants to optimize the model using automatic model tuning. How can he decide which hyper-parameter to be optimize and the range to be used?**

A He has can choose any of the parameter and he has to specify the range

B He has can choose any of the parameter and range will be automatically detected

C He has to consult the documentation regarding the parameter and range will be automatically detected

D He has to consult the documentation regarding the parameter and he has to specify the range

7. **You want to use the model artifacts from another training job. So that you don't want to start from scratch. Which type of training would help in this case?**

   A Split training
   B Artifact training
   C Prior training
   D Incremental training

8. **You want to create an agent to play the game of GO, that is able to learn from playing and it should be able to beat human players. What type of a problem it is?**

   A Linear regression problem
   B Reinforcement problem
   C Classification problem
   D Object detection problem

9. **You need to cluster data into 4 different groups based on similarity. You don't have any label associated to the data. What algorithm can help you?**

   A K-Nearest Neighbour
   B Classification
   C K-Means
   D XGBoost

10. **You are given a large set of unstructured text data and you want to perform sentiment analysis on the data. What could be the required pre-processing step?**

    A Use Stemming
    B Use One-hot encoding
    C Use Categorical encoding
    D Use vector transformation

9. **You want to build an image classification algorithm using Sagemaker. What are the steps involved in creating an image classifier?**

   A Upload the dataset to RDF and use built-in Sagemaker Image Classifier

   B Upload the dataset to S3 and use built-in Sagemaker XGBoost

   C Upload the dataset to S3 and use built-in Sagemaker Image Classifier

   D Upload the dataset to RDF and use built-in Sagemaker XGBoost

10. **You are given access to a large dataset containing about million rows of data. The dataset contains information about customer feedback on different products. You are tasked to build a recommendation model that is able to predict how many people would like the new product they are launching. There are some missing data in the dataset making it a sparse dataset. What algorithm can be used?**

    A Principal Component Analysis

    B K-Means

    C Seq2Seq

    D Factorization Machines

ANSWER KEY:

1. A	2. C	3. C	4. B	5. A, C
6. D	7. D	8. B	9. C	10. A