vmexam.com
#1 Cloud and VM Certification Guide

AWS MLS-C01

AWS MACHINE LEARNING SPECIALTY CERTIFICATION QUESTIONS & ANSWERS

Exam Summary – Syllabus – Questions

MLS-C01

AWS Certified Machine Learning - Specialty
65 Questions Exam – 750 / 1000 Cut Score – Duration of 180 minutes

www.VMExam.com

Table of Contents

Know Your MLS-C01 Certification Well:

The MLS-C01 is best suitable for candidates who want to gain knowledge in the AWS Specialty. Before you start your MLS-C01 preparation you may struggle to get all the crucial Machine Learning Specialty materials like MLS-C01 syllabus, sample questions, study guide.

But don't worry the MLS-C01 PDF is here to help you prepare in a stress free manner.

The PDF is a combination of all your queries like-

- What is in the MLS-C01 syllabus?
- How many questions are there in the MLS-C01 exam?
- Which Practice test would help me to pass the MLS-C01 exam at the first attempt?

Passing the MLS-C01 exam makes you AWS Certified Machine Learning - Specialty. Having the Machine Learning Specialty certification opens multiple opportunities for you. You can grab a new job, get a higher salary or simply get recognition within your current organization.

AWS MLS-C01 Machine Learning Specialty Certification Details:

Exam Name	AWS Certified Machine Learning - Specialty (Machine Learning Specialty)
Exam Code	MLS-C01
Exam Price	$300 USD
Duration	180 minutes
Number of Questions	65
Passing Score	750 / 1000
Recommended Training / Books	**Practical Data Science with Amazon SageMaker** **The Machine Learning Pipeline on AWS** **Deep Learning on AWS**
Schedule Exam	**PEARSON VUE**
Sample Questions	**AWS MLS-C01 Sample Questions**
Recommended Practice	**AWS Certified Machine Learning - Specialty Practice Test**

MLS-C01 Syllabus:

Section	Objectives
Data Engineering - 20%	
Create data repositories for machine learning.	- Identify data sources (e.g., content and location, primary sources such as user data) - Determine storage mediums (e.g., DB, Data Lake, S3, EFS, EBS)
Identify and implement a data ingestion solution.	- Data job styles/types (batch load, streaming) - Data ingestion pipelines (Batch-based ML workloads and streaming-based ML workloads) • Kinesis • Kinesis Analytics • Kinesis Firehose • EMR • Glue - Job scheduling
Identify and implement a data transformation solution.	- Transforming data transit (ETL: Glue, EMR, AWS Batch) - Handle ML-specific data using map reduce (Hadoop, Spark, Hive)
Exploratory Data Analysis - 24%	
Sanitize and prepare data for modeling.	- Identify and handle missing data, corrupt data, stop words, etc. - Formatting, normalizing, augmenting, and scaling data - Labeled data (recognizing when you have enough labeled data and identifying mitigation strategies [Data labeling tools (Mechanical Turk, manual labor)])
Perform feature engineering.	- Identify and extract features from data sets, including from data sources such as text, speech, image, public datasets, etc. - Analyze/evaluate feature engineering concepts (binning, tokenization, outliers, synthetic features, 1 hot encoding, reducing dimensionality of data)
Analyze and visualize data for machine learning.	- Graphing (scatter plot, time series, histogram, box plot) - Interpreting descriptive statistics (correlation, summary statistics, p value) - Clustering (hierarchical, diagnosing, elbow plot, cluster size)

Section	Objectives
Modeling - 36%	
Frame business problems as machine learning problems.	- Determine when to use/when not to use ML - Know the difference between supervised and unsupervised learning - Selecting from among classification, regression, forecasting, clustering, recommendation, etc.
Select the appropriate model(s) for a given machine learning problem.	- Xgboost, logistic regression, K-means, linear regression, decision trees, random forests, RNN, CNN, Ensemble, Transfer learning - Express intuition behind models
Train machine learning models.	- Train validation test split, cross-validation - Optimizer, gradient descent, loss functions, local minima, convergence, batches, probability, etc. - Compute choice (GPU vs. CPU, distributed vs. non-distributed, platform [Spark vs. non-Spark]) - Model updates and retraining • Batch vs. real-time/online
Perform hyperparameter optimization.	- Regularization • Drop out • L1/L2 - Cross validation - Model initialization - Neural network architecture (layers/nodes), learning rate, activation functions - Tree-based models (# of trees, # of levels) - Linear models (learning rate)
Evaluate machine learning models.	- Avoid overfitting/underfitting (detect and handle bias and variance) - Metrics (AUC-ROC, accuracy, precision, recall, RMSE, F1 score) - Confusion matrix - Offline and online model evaluation, A/B testing - Compare models using metrics (time to train a model, quality of model, engineering costs) - Cross validation
Machine Learning Implementation and Operations - 20%	
Build machine learning solutions for performance,	- AWS environment logging and monitoring • CloudTrail and CloudWatch • Build error monitoring

Section	Objectives
availability, scalability, resiliency, and fault tolerance.	- Multiple regions, Multiple AZs - AMI/golden image - Docker containers - Auto Scaling groups - Rightsizing • Instances • Provisioned IOPS • Volumes - Load balancing - AWS best practices
Recommend and implement the appropriate machine learning services and features for a given problem.	- ML on AWS (application services) • Poly • Lex • Transcribe - AWS service limits - Build your own model vs. SageMaker built-in algorithms - Infrastructure: (spot, instance types), cost considerations • Using spot instances to train deep learning models using AWS Batch
Apply basic AWS security practices to machine learning solutions.	- IAM - S3 bucket policies - Security groups - VPC - Encryption/anonymization
Deploy and operationalize machine learning solutions.	- Exposing endpoints and interacting with them - ML model versioning - A/B testing - Retrain pipelines - ML debugging/troubleshooting • Detect and mitigate drop in performance • Monitor performance of the model

AWS MLS-C01 Sample Questions:

Question: 1

A Data Scientist is evaluating different binary classification models. A false positive result is 5 times more expensive (from a business perspective) than a false negative result.

The models should be evaluated based on the following criteria:

1) Must have a recall rate of at least 80%
2) Must have a false positive rate of 10% or less
3) Must minimize business costs

After creating each binary classification model, the Data Scientist generates the corresponding confusion matrix.

Which confusion matrix represents the model that satisfies the requirements?

a) TN = 91, FP = 9 FN = 22, TP = 78
b) TN = 99, FP = 1 FN = 21, TP = 79
c) TN = 96, FP = 4 FN = 10, TP = 90
d) TN = 98, FP = 2 FN = 18, TP = 82

Answer: d

Question: 2

A Data Scientist uses logistic regression to build a fraud detection model. While the model accuracy is 99%, 90% of the fraud cases are not detected by the model.

What action will definitively help the model detect more than 10% of fraud cases?

a) Using oversampling to balance the dataset
b) Using regularization to reduce overfitting
c) Decreasing the class probability threshold
d) Using undersampling to balance the dataset

Answer: c

Question: 3

A Data Scientist is working on optimizing a model during the training process by varying multiple parameters. The Data Scientist observes that, during multiple runs with identical parameters, the loss function converges to different, yet stable, values.

What should the Data Scientist do to improve the training process?

a) Increase the learning rate. Keep the batch size the same.
b) Reduce the batch size. Decrease the learning rate.
c) Keep the batch size the same. Decrease the learning rate.
d) Do not change the learning rate. Increase the batch size.

Answer: b

Question: 4

A company is setting up a system to manage all of the datasets it stores in Amazon S3.

The company would like to automate running transformation jobs on the data and maintaining a catalog of the metadata concerning the datasets. The solution should require the least amount of setup and maintenance.

Which solution will allow the company to achieve its goals?

a) Create an Amazon EMR cluster with Apache Hive installed. Then, create a Hive metastore and a script to run transformation jobs on a schedule.
b) Create an AWS Glue crawler to populate the AWS Glue Data Catalog. Then, author an AWS Glue ETL job, and set up a schedule for data transformation jobs.
c) Create an Amazon EMR cluster with Apache Spark installed. Then, create an Apache Hive metastore and a script to run transformation jobs on a schedule.
d) Create an AWS Data Pipeline that transforms the data. Then, create an Apache Hive metastore and a script to run transformation jobs on a schedule.

Answer: b

Question: 5

An insurance company needs to automate claim compliance reviews because human reviews are expensive and error-prone. The company has a large set of claims and a compliance label for each.

Each claim consists of a few sentences in English, many of which contain complex related information. Management would like to use Amazon SageMaker built-in algorithms to design a machine learning supervised model that can be trained to read each claim and predict if the claim is compliant or not.

Which approach should be used to extract features from the claims to be used as inputs for the downstream supervised task?

a) Derive a dictionary of tokens from claims in the entire dataset. Apply one-hot encoding to tokens found in each claim of the training set. Send the derived features space as inputs to an Amazon SageMaker builtin supervised learning algorithm.
b) Apply Amazon SageMaker BlazingText in Word2Vec mode to claims in the training set. Send the derived features space as inputs for the downstream supervised task.
c) Apply Amazon SageMaker BlazingText in classification mode to labeled claims in the training set to derive features for the claims that correspond to the compliant and non-compliant labels, respectively.
d) Apply Amazon SageMaker Object2Vec to claims in the training set. Send the derived features space as inputs for the downstream supervised task.

Answer: d

Question: 6

A Machine Learning team has several large CSV datasets in Amazon S3. Historically, models built with the Amazon SageMaker Linear Learner algorithm have taken hours to train on similar-sized datasets. The team's leaders need to accelerate the training process.

What can a Machine Learning Specialist do to address this concern?

a) Use Amazon SageMaker Pipe mode.
b) Use Amazon Machine Learning to train the models.
c) Use Amazon Kinesis to stream the data to Amazon SageMaker.
d) Use AWS Glue to transform the CSV dataset to the JSON format.

Answer: a

Question: 7

A Machine Learning Engineer is preparing a data frame for a supervised learning task with the Amazon SageMaker Linear Learner algorithm.

The ML Engineer notices the target label classes are highly imbalanced and multiple feature columns contain missing values. The proportion of missing values across the entire data frame is less than 5%.

What should the ML Engineer do to minimize bias due to missing values?

a) Replace each missing value by the mean or median across non-missing values in same row.
b) Delete observations that contain missing values because these represent less than 5% of the data.
c) Replace each missing value by the mean or median across non-missing values in the same column.
d) For each feature, approximate the missing values using supervised learning based on other features.

Answer: d

Question: 8

A term frequency–inverse document frequency (tf–idf) matrix using both unigrams and bigrams is built from a text corpus consisting of the following two sentences:

1. Please call the number below.
2. Please do not call us.

What are the dimensions of the tf–idf matrix?

a) (2, 16)
b) (2, 8)
c) (2, 10)
d) (8, 10)

Answer: a

Question: 9

A company has collected customer comments on its products, rating them as safe or unsafe, using decision trees. The training dataset has the following features:

id, date, full review, full review summary, and a binary safe/unsafe tag. During training, any data sample with missing features was dropped. In a few instances, the test set was found to be missing the full review text field.

For this use case, which is the most effective course of action to address test data samples with missing features?

a) Drop the test samples with missing full review text fields, and then run through the test set.
b) Copy the summary text fields and use them to fill in the missing full review text fields, and then run through the test set.
c) Use an algorithm that handles missing data better than decision trees.
d) Generate synthetic data to fill in the fields that are missing data, and then run through the test set.

Answer: b

Question: 10

A company is interested in building a fraud detection model. Currently, the Data Scientist does not have a sufficient amount of information due to the low number of fraud cases.

Which method is MOST likely to detect the GREATEST number of valid fraud cases?

a) Oversampling using bootstrapping
b) Undersampling
c) Oversampling using SMOTE
d) Class weight adjustment

Answer: c

Study Guide to Crack AWS Machine Learning Specialty MLS-C01 Exam:

- Getting details of the MLS-C01 syllabus, is the first step of a study plan. This pdf is going to be of ultimate help. Completion of the syllabus is must to pass the MLS-C01 exam.
- Making a schedule is vital. A structured method of preparation leads to success. A candidate must plan his schedule and follow it rigorously to attain success.
- Joining the AWS provided training for MLS-C01 exam could be of much help. If there is specific training for the exam, you can discover it from the link above.
- Read from the MLS-C01 sample questions to gain your idea about the actual exam questions. In this PDF useful sample questions are provided to make your exam preparation easy.
- Practicing on MLS-C01 practice tests is must. Continuous practice will make you an expert in all syllabus areas.

Reliable Online Practice Test for MLS-C01 Certification

Make VMExam.com your best friend during your AWS Certified Machine Learning - Specialty exam preparation. We provide authentic practice tests for the MLS-C01 exam. Experts design these online practice tests, so we can offer you an exclusive experience of taking the actual MLS-C01 exam. We guarantee you 100% success in your first exam attempt if you continue practicing regularly. Don't bother if you don't get 100% marks in initial practice exam attempts. Just utilize the result section to know your strengths and weaknesses and prepare according to that until you get 100% with our practice tests. Our evaluation makes you confident, and you can score high in the MLS-C01 exam.

Start Online practice of MLS-C01 Exam by visiting URL

https://www.vmexam.com/aws/mls-c01-aws-certified-machine-learning-specialty