AWS CERTIFIED MACHINE LEARNING SPECIALTY (MLS-C01) PRACTICE TEST (SAMPLE)

STUDY GUIDE



SAMPLE STUDY GUIDE WITH LIMITED QUESTIONS

VISIT US ONLINE FOR THE FULL VERSION STUDY GUIDE

https://awscertmachinelearningspecialty.examzify.com

Copyright © 2026 by Examzify - A Kaluba Technologies Inc. product.

ALL RIGHTS RESERVED.

No part of this book may be reproduced or transferred in any form or by any means, graphic, electronic, or mechanical, including photocopying, recording, web distribution, taping, or by any information storage retrieval system, without the written permission of the author.

Notice: Examzify makes every reasonable effort to obtain accurate, complete, and timely information about this product from reliable sources.

SAMPLE

Table of Contents

Copyright 1

Table of Contents 2

Introduction 3

How to Use This Guide 4

Questions 5

Answers 8

Explanations 10

Next Steps 16

SAMPLE

Introduction

Preparing for a certification exam can feel overwhelming, but with the right tools, it becomes an opportunity to build confidence, sharpen your skills, and move one step closer to your goals. At Examzify, we believe that effective exam preparation isn't just about memorization, it's about understanding the material, identifying knowledge gaps, and building the test-taking strategies that lead to success.

This guide was designed to help you do exactly that.

Whether you're preparing for a licensing exam, professional certification, or entry-level qualification, this book offers structured practice to reinforce key concepts. You'll find a wide range of multiple-choice questions, each followed by clear explanations to help you understand not just the right answer, but why it's correct.

The content in this guide is based on real-world exam objectives and aligned with the types of questions and topics commonly found on official tests. It's ideal for learners who want to:

- Practice answering questions under realistic conditions,
- Improve accuracy and speed,
- Review explanations to strengthen weak areas, and
- Approach the exam with greater confidence.

We recommend using this book not as a stand-alone study tool, but alongside other resources like flashcards, textbooks, or hands-on training. For best results, we recommend working through each question, reflecting on the explanation provided, and revisiting the topics that challenge you most.

Remember: successful test preparation isn't about getting every question right the first time, it's about learning from your mistakes and improving over time. Stay focused, trust the process, and know that every page you turn brings you closer to success.

Let's begin.

SAMPLE

How to Use This Guide

This guide is designed to help you study more effectively and approach your exam with confidence. Whether you're reviewing for the first time or doing a final refresh, here's how to get the most out of your Examzify study guide:

1. Start with a Diagnostic Review

Skim through the questions to get a sense of what you know and what you need to focus on. Your goal is to identify knowledge gaps early.

2. Study in Short, Focused Sessions

Break your study time into manageable blocks (e.g. 30 – 45 minutes). Review a handful of questions, reflect on the explanations.

3. Learn from the Explanations

After answering a question, always read the explanation, even if you got it right. It reinforces key points, corrects misunderstandings, and teaches subtle distinctions between similar answers.

4. Track Your Progress

Use bookmarks or notes (if reading digitally) to mark difficult questions. Revisit these regularly and track improvements over time.

5. Simulate the Real Exam

Once you're comfortable, try taking a full set of questions without pausing. Set a timer and simulate test-day conditions to build confidence and time management skills.

6. Repeat and Review

Don't just study once, repetition builds retention. Re-attempt questions after a few days and revisit explanations to reinforce learning. Pair this guide with other Examzify tools like flashcards, and digital practice tests to strengthen your preparation across formats.

There's no single right way to study, but consistent, thoughtful effort always wins. Use this guide flexibly, adapt the tips above to fit your pace and learning style. You've got this!

SAMPLE

Questions

SAMPLE

1. **What service provides an on-demand workforce for tasks like recognizing objects in photographs?**

   A. Amazon Mechanical Turk

   B. Amazon SageMaker

   C. Amazon Rekognition

   D. Amazon Transcribe

2. **What is the AWS managed container image registry service that is secure, scalable, and reliable?**

   A. Amazon Elastic Container Service

   B. Amazon Elastic Kubernetes Service

   C. Amazon Elastic Container Registry

   D. Amazon Fargate

3. **Which technique is used to convert a categorical feature into a binary format, with each category represented as a separate binary feature?**

   A. Label Encoding

   B. Ordinal Encoding

   C. One-hot encoding

   D. Binary Encoding

4. **Which evaluation method is best for a model that shows a significant difference between training and validation accuracies?**

   A. Cross-Validation

   B. Grid Search

   C. Data Augmentation

   D. Gated Validation

5. **What is a primary benefit of using Amazon Kinesis Data Analytics applications?**

   A. To visualize data in real-time

   B. To implement machine learning models

   C. To process and analyze streaming data

   D. To store large datasets

SAMPLE

6. What is the goal of feature scaling in machine learning?

A. To select features for the model

B. To convert features into categorical values

C. To ensure all features contribute equally

D. To reduce model complexity

7. What enables private connections between your VPC and supported AWS services powered by AWS PrivateLink?

A. VPC Peering

B. VPC Endpoint

C. AWS Direct Connect

D. AWS Transit Gateway

8. What service automatically handles server and operating system maintenance, capacity provisioning, and scaling?

A. AWS Glue

B. AWS Lambda

C. Amazon EC2

D. AWS Batch

9. What is the main function of the K-Means clustering algorithm?

A. To classify data using labeled examples

B. To partition data into distinct clusters

C. To reduce dimensionality of data

D. To optimize hyperparameters in models

10. Which Amazon service uses deep learning technology to analyze images and videos?

A. Amazon SageMaker

B. Amazon Rekognition

C. Amazon Mechanical Turk

D. Amazon Fraud Detector

SAMPLE

Answers

SAMPLE

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

SAMPLE

Explanations

SAMPLE

1. What service provides an on-demand workforce for tasks like recognizing objects in photographs?

A. Amazon Mechanical Turk

B. Amazon SageMaker

C. Amazon Rekognition

D. Amazon Transcribe

Amazon Mechanical Turk provides an on-demand workforce that is ideal for tasks requiring human intelligence, such as recognizing objects in photographs. This service enables users to create Human Intelligence Tasks (HITs), where workers can perform various tasks, including image labeling and object recognition. This flexibility makes it particularly useful for machine learning projects that require annotated data for training. While Amazon Rekognition is a powerful service for image and video analysis, it is an automated solution that uses machine learning to identify objects without human intervention. Thus, it does not directly provide a workforce for human-centric tasks. Amazon SageMaker is a comprehensive machine learning service, but it does not directly connect users to a workforce for manual tasks. Amazon Transcribe focuses solely on converting speech into text and is not relevant to visual recognition tasks. Therefore, the appropriate answer for an on-demand workforce suitable for recognizing objects in photographs is Amazon Mechanical Turk.

2. What is the AWS managed container image registry service that is secure, scalable, and reliable?

A. Amazon Elastic Container Service

B. Amazon Elastic Kubernetes Service

C. Amazon Elastic Container Registry

D. Amazon Fargate

The AWS managed container image registry service that is secure, scalable, and reliable is Amazon Elastic Container Registry (ECR). ECR enables you to store, manage, and deploy Docker container images. Being a fully managed service, it eliminates the need to operate your own container repositories, simplifying the process of container management. Amazon ECR is designed with security features such as integrated IAM policies for fine-grained access control, making it suitable for private registries. It supports image lifecycle policies, allowing efficient management and cleanup of images, which enhances its scalability. ECR also seamlessly integrates with other AWS services and developer tools, facilitating CI/CD workflows. In contrast, the other options listed serve different purposes within the AWS ecosystem. Amazon Elastic Container Service (ECS) is a container orchestration service that allows you to run and manage containerized applications, while Amazon Elastic Kubernetes Service (EKS) provides a managed Kubernetes platform. Amazon Fargate, on the other hand, is a serverless compute engine for containers that fits with ECS and EKS by allowing you to run containers without managing the underlying servers. Each of these services complements ECR but does not provide the specific function of managing container images.

SAMPLE

3. Which technique is used to convert a categorical feature into a binary format, with each category represented as a separate binary feature?

A. Label Encoding

B. Ordinal Encoding

C. One-hot encoding

D. Binary Encoding

One-hot encoding is the technique used to convert a categorical feature into a binary format, where each category is represented as a separate binary feature. This method creates new binary columns for each unique category in the original feature, where each column indicates the presence (1) or absence (0) of that category for a given observation. For instance, if you have a categorical feature "Color" with the categories "Red," "Green," and "Blue," one-hot encoding will create three new columns: one for Red, one for Green, and one for Blue. Each row will have a value of 1 in one of these columns and 0s in the others, effectively transforming the single categorical feature into three binary features. This technique is particularly useful in machine learning algorithms that require numerical input, ensuring that the model can learn patterns without implying any ordinal relationship between the categories, which is not the case with categorical data.

4. Which evaluation method is best for a model that shows a significant difference between training and validation accuracies?

A. Cross-Validation

B. Grid Search

C. Data Augmentation

D. Gated Validation

Cross-validation is the best evaluation method for a model that demonstrates a significant discrepancy between training and validation accuracies. This difference typically indicates that the model may be overfitting the training data, capturing noise rather than the underlying patterns. Cross-validation helps mitigate this issue by systematically partitioning the training dataset into multiple subsets. The model is then trained on a portion of the data and validated on the remaining part in several iterations. This process allows for a more reliable estimation of the model's performance on unseen data, as it evaluates the model's accuracy across different subsets instead of relying solely on a single validation split. In scenarios where there is notable overfitting, cross-validation can help in identifying if the model consistently performs poorly on validation sets across different data partitions, thereby providing insights into whether adjustments, such as regularization or model architecture modifications, are necessary. The other methods listed, like grid search, are mostly focused on hyperparameter tuning, and while they can improve model performance, they do not specifically address the fundamental issue of overfitting. Data augmentation is useful for improving model robustness by increasing the training dataset size and diversity, but it doesn't directly evaluate model discrepancies. Gated validation is not a standard term in model evaluation and does not apply

SAMPLE

5. What is a primary benefit of using Amazon Kinesis Data Analytics applications?

A. To visualize data in real-time

B. To implement machine learning models

C. To process and analyze streaming data

D. To store large datasets

Amazon Kinesis Data Analytics is designed specifically for processing and analyzing streaming data in real-time. This allows organizations to gain immediate insights from their data as it flows through the system, enabling timely decision-making and action based on the most current information available. The application transforms and analyzes streaming data using standard SQL queries, making it accessible and manageable. This capability is crucial for various use cases, such as monitoring streaming data from IoT devices, processing logs in real-time, or conducting fraud detection as transactions occur. By focusing on the ability to process and analyze data as it arrives, Kinesis Data Analytics empowers organizations to turn raw data into actionable insights almost instantaneously. The other options, while they touch on aspects relevant to data handling, do not fully encapsulate the primary function of Kinesis Data Analytics. Visualizing data in real-time and implementing machine learning models may utilize the outputs of Kinesis, but they are not the core feature of the service itself. Storing large datasets, on the other hand, is not a function of Kinesis Data Analytics; its strength lies in processing data rather than storing it, which is typically handled by other Amazon services like Amazon S3 or Amazon Redshift.

6. What is the goal of feature scaling in machine learning?

A. To select features for the model

B. To convert features into categorical values

C. To ensure all features contribute equally

D. To reduce model complexity

Feature scaling aims to ensure that all features contribute equally to the model's performance. In many machine learning algorithms, particularly those sensitive to the scale of the input features, such as gradient descent-based methods, the scale of the features can significantly impact model training. When features have different ranges or units, models may pay more attention to the larger-scaled features, leading to biased results. By performing feature scaling, all features are normalized or standardized to a uniform scale. This can be achieved through techniques such as min-max scaling, which transforms features to a range between 0 and 1, or Z-score normalization, which standardizes the features to have a mean of 0 and standard deviation of 1. This treatment enhances model convergence rates and guarantees that the algorithm learns from the data more effectively, improving both performance and accuracy. The other options do not directly relate to the fundamental aim of feature scaling. Selecting features is a separate process of feature selection, while converting features into categorical values addresses a different aspect of data preprocessing. Reducing model complexity pertains to other techniques such as regularization, rather than focusing on the scaling of feature inputs.

SAMPLE

7. What enables private connections between your VPC and supported AWS services powered by AWS PrivateLink?

A. VPC Peering

B. VPC Endpoint

C. AWS Direct Connect

D. AWS Transit Gateway

The correct answer is VPC Endpoint, as it is designed specifically to enable private connections between your Virtual Private Cloud (VPC) and supported AWS services without requiring public IPs or exposing your traffic to the public internet. When you create a VPC Endpoint, you can privately connect your VPC to various AWS services, enhancing security and minimizing the risk of data exposure. VPC Endpoints come in two types: Gateway Endpoints for specific AWS services like S3 and DynamoDB, and Interface Endpoints that allow connections to other AWS services and your own applications hosted in a different VPC. By using VPC Endpoints, you can ensure that the communication between your VPC and the services happens over the AWS network, providing a more secure and reliable way to access these services. The other choices do not specifically facilitate private connections in this context. VPC Peering establishes a direct network connection between two VPCs, but not directly with AWS services. AWS Direct Connect provides a dedicated connection between your on-premises data center and AWS but does not create private connections to AWS services within your VPC. AWS Transit Gateway simplifies and centralizes your network connections across multiple VPCs and on-premises environments but does not directly relate to accessing supported

8. What service automatically handles server and operating system maintenance, capacity provisioning, and scaling?

A. AWS Glue

B. AWS Lambda

C. Amazon EC2

D. AWS Batch

The service that automatically handles server and operating system maintenance, capacity provisioning, and scaling is AWS Lambda. This serverless computing service allows users to run code without worrying about the underlying infrastructure. It automatically manages the compute resources required to run your application, including provisioning and scaling as needed. With AWS Lambda, developers can focus solely on their code instead of the operational aspects like server maintenance or capacity management. This is particularly useful for applications with variable workloads, as Lambda can scale up or down seamlessly to meet demand without manual intervention. The other options provided serve different purposes: AWS Glue is designed for data integration and ETL processes, Amazon EC2 offers virtual servers but requires management of instances and scaling by the user, and AWS Batch is tailored for running batch processing jobs but does not provide the same level of abstraction from infrastructure management as Lambda does.

SAMPLE

9. What is the main function of the K-Means clustering algorithm?

A. To classify data using labeled examples

B. To partition data into distinct clusters

C. To reduce dimensionality of data

D. To optimize hyperparameters in models

The K-Means clustering algorithm primarily serves the function of partitioning data into distinct clusters. It operates by grouping a set of data points into a predefined number of clusters, often denoted as 'k'. Each cluster is characterized by its centroid, which is the average of all data points belonging to that cluster. The algorithm iteratively adjusts these centroids and the association of data points to clusters to achieve a partitioning that minimizes the internal variance within each cluster, effectively making similar data points more closely grouped together. K-Means is particularly useful in exploratory data analysis, allowing for the identification of underlying patterns and structures within unlabelled data. The simple yet effective approach makes it ideal for scenarios where the primary goal is to identify natural groupings in the dataset without any prior label information. The other options represent different machine learning functions: classifying using labeled examples is the domain of supervised learning; dimensionality reduction involves techniques like PCA (Principal Component Analysis) and is used to decrease the number of features while retaining important information; and hyperparameter optimization pertains to the tuning of model parameters to improve performance on datasets. Each of these functions addresses specific challenges in machine learning, but they do not align with the primary goal of K-Means clustering.

10. Which Amazon service uses deep learning technology to analyze images and videos?

A. Amazon SageMaker

B. Amazon Rekognition

C. Amazon Mechanical Turk

D. Amazon Fraud Detector

Amazon Rekognition is the service designed specifically for analyzing images and videos using deep learning technology. It offers a range of features such as object and scene detection, facial analysis, facial recognition, and activity detection. By leveraging convolutional neural networks (CNNs) and other deep learning techniques, Rekognition can process large volumes of media efficiently and accurately to extract meaningful insights. This service is particularly useful for developers who need to automate image and video analysis tasks, such as identifying and labeling objects in an image or conducting security checks through facial recognition. The ability to utilize pre-trained models makes it accessible to users who may not have extensive machine learning expertise. In contrast, Amazon SageMaker is a more general service that provides a comprehensive environment for developing, training, and deploying machine learning models, but it does not specialize explicitly in image and video analysis. Amazon Mechanical Turk focuses on crowdsourcing human intelligence tasks and does not involve deep learning analysis features directly. Amazon Fraud Detector is primarily aimed at detecting fraudulent activity in online transactions and is not tailored specifically for media analysis.

SAMPLE

Next Steps

Congratulations on reaching the final section of this guide. You've taken a meaningful step toward passing your certification exam and advancing your career.

As you continue preparing, remember that consistent practice, review, and self-reflection are key to success. Make time to revisit difficult topics, simulate exam conditions, and track your progress along the way.

If you need help, have suggestions, or want to share feedback, we'd love to hear from you. Reach out to our team at hello@examzify.com.

Or visit your dedicated course page for more study tools and resources:

https://awscertmachinelearningspecialty.examzify.com

We wish you the very best on your exam journey. You've got this!

SAMPLE