

AWS Certified Machine Learning - Specialty (MLS-C01) 考試指南

簡介

AWS Certified Machine Learning - Specialty (MLS-C01) 考試專為執行人工智慧和機器學習 (AI/ML) 開發，或擔任資料科學角色的個人設計。本考試旨在驗證考生是否有能力使用 AWS 雲端，針對特定商業問題設計、建置、部署、最佳化、訓練、調校和維護 ML 解決方案。

本考試還會驗證考生是否有能力完成以下任務：

- 為指定的商業問題選擇適當的機器學習 (ML) 方法，並證明其合理性。
- 找出適合實作機器學習 (ML) 解決方案的 AWS 服務。
- 設計和實作可擴展、成本最佳化、可靠和安全的機器學習 (ML) 解決方案。

目標考生描述

目標考生應具備 2 年以上在 AWS Cloud 開發、架構和執行 ML 或深度學習工作負載的經驗。

AWS 知識推薦

目標考生應具備以下 AWS 知識：

- 表達基本機器學習 (ML) 演算法直觀原理的能力
- 執行基本超參數最佳化的經驗
- ML 和深度學習架構的經驗
- 遵循模型訓練最佳實務的能力
- 遵循部署最佳實務的能力
- 遵循營運最佳實務的能力

超出目標考生範圍的知識

下列清單包含目標考生預計未擁有的知識。此表並非詳盡無遺。以下領域的知識超出考試範圍：

- 廣泛或複雜的演算法開發
- 廣泛的超參數最佳化
- 複雜的數學證明與運算
- 進階連網與網路設計
- 進階資料庫、安全性與 **DevOps** 概念
- **Amazon EMR** 的 **DevOps** 相關任務

如需考試中可能出現的技術和概念清單、考試範圍內的 **AWS** 服務和功能清單，以及超出範圍的 **AWS** 服務和功能清單，請參閱附錄。

考試內容

答案類型

本考試含有兩種題型：

- **單選題**：每題有一個正確的答案和三個不正確的答案 (干擾選項)
- **複選題**：每題有五個以上的答案選項，其中有兩個以上的正確答案

選擇一或多個最能完整說明或回答問題的答案。干擾選項或不正確的答案是不具備完整知識或技能的考生可能會選擇的答案選項。干擾選項符合內容範圍，但實際上是似是而非的答案。

未回答的問題將以不正確計算；猜答不會扣分。本考試設有 **50** 個會影響您分數的問題。

不計分內容

本考試設有 **15** 個不會影響您分數的不計分問題。**AWS** 會透過這些不計分問題來收集與考生表現相關的資訊，藉以評估這些問題能否在將來作為計分問題使用。考試中不能識別這些不計分的問題。

考試結果

AWS Certified Machine Learning - Specialty (MLS-C01) 考試結果分為合格和不合格兩種。本考試的評分，將按照 **AWS** 專業人員根據認證產業最佳實務和指南所制定的最低標準來進行。

您的考試結果將以 **100** 至 **1,000** 分的比例分數呈現。最低及格分數為 **750** 分。您的分數表示您在考試中的整體表現，以及您是否通過考試。比例分數模型有助將多種考試類型的得分放在一起比較，這些考試的難度級別可能略有不同。

您的分數報告中可能會包含您在各個部分級別的表現分級表。本考試採用補償性評分模式，這代表您並不需要在每個部分都獲得合格分數。您只需要通過整體的考試。

考試的每個部分都有一個特定的權重，因此有些部分需回答的問題比其他部分多。分級表包含一般資訊，其中重點標示了您的優勢和劣勢。在解釋部分級別意見時請格外慎重。

命題大綱

此考試指南包括權重、測試領域和考試的任務說明。本指南並不提供考試內容的詳盡列表。但是，每個任務說明的其他上下文均有助於您準備考試。

考試具有以下測試領域和權重：

- 領域 1：資料工程 (佔計分內容的 20%)
- 領域 2：探索資料分析 (佔計分內容的 24%)
- 領域 3：建模 (佔計分內容的 36%)
- 領域 4：機器學習實作和操作 (佔計分內容的 20%)

領域 1：資料工程

任務說明 1.1：建立 ML 的資料儲存庫。

- 識別資料來源 (例如內容和位置、使用者資料等主要來源)。
- 確定儲存媒介 (例如資料庫、Amazon S3、Amazon Elastic File System [Amazon EFS]、Amazon Elastic Block Store [Amazon EBS])。

任務說明 1.2：識別和實作資料擷取解決方案。

- 識別資料任務樣式和任務類型 (例如批次載入、串流)。
- 協調資料擷取管道 (批次式 ML 工作負載和串流式 ML 工作負載)。
 - Amazon Kinesis
 - Amazon Data Firehose
 - Amazon EMR
 - AWS Glue
 - Amazon Managed Service for Apache Flink
- 排程任務。

任務說明 1.3：識別和實作資料轉換解決方案。

- 轉換傳輸中的資料 (ETL、AWS Glue、Amazon EMR、AWS Batch)。
- 利用 MapReduce (例如 Apache Hadoop、Apache Spark、Apache Hive) 處理特定於 ML 的資料。

領域 2：探索資料分析

任務說明 2.1：淨化和準備資料以供建模使用。

- 識別和處理遺失的資料、損壞的資料和停用詞。
- 格式化、正規化、擴充和擴展資料。
- 確定是否有足夠的標籤資料。
 - 識別緩解策略。
 - 使用資料標籤工具 (如 **Amazon Mechanical Turk**)。

任務說明 2.2：執行特徵工程。

- 識別和擷取資料集中的功能，包括文字、語音、映像和公開資料集等資料來源。
- 分析和評估特徵工程概念 (例如分箱、記號化、異常值、合成特徵、獨熱編碼、減少資料的維度)。

任務說明 2.3：分析 **ML** 資料並以視覺化方式呈現。

- 建立圖形 (例如散佈圖、時間序列、長條圖、箱型圖)。
- 解釋描述性統計值 (例如關聯性、統計摘要、**p** 值)。
- 執行叢集分析 (例如階層式、診斷、肘形圖、叢集大小)。

領域 3：建模

任務說明 3.1：將商業問題套入 **ML** 問題框架。

- 確定使用或不使用 **ML** 的時機。
- 瞭解監督學習和無監督學習之間的區別。
- 從分類、回歸、預測、叢集、推薦和基礎模型中選取。

任務說明 3.2：針對特定 **ML** 問題選取適當的模型。

- **XGBoost**、邏輯迴歸、**k-means**、線性迴歸、決策樹、隨機森林、**RNN**、**CNN**、組合、遷移學習以及大型語言模型 (**LLM**)
- 表達模型背後的直覺。

任務說明 3.3：訓練 ML 模型。

- 在訓練和驗證之間分割資料 (例如交叉驗證)。
- 瞭解 ML 訓練的最佳化技術 (例如梯度下降法、損失函數、收斂)。
- 選擇適當的運算資源 (例如 GPU 或 CPU、分散式或非分散式)。
 - 選擇適當的運算平台 (Spark 或非 SPARK)。
- 更新和重新訓練模型。
 - 批次或即時/線上

任務說明 3.4：執行超參數最佳化。

- 執行正規化。
 - 退出
 - L1/L2
- 執行交叉驗證。
- 初始化模型。
- 瞭解神經網路架構 (層與節點)、學習速率和啟動函數。
- 瞭解以決策樹為基礎的模型 (決策樹數量、層數)。
- 瞭解線性模型 (學習速率)。

任務說明 3.5：評估 ML 模型。

- 避免過度擬合或低度擬合。
 - 偵測和處理偏差和變異。
- 評估指標 (例如：曲線下的面積 [AUC]-接收者操作特徵 [ROC]、準確性、精確度、召回率、均方根誤差 [RMSE]、F1 分數)。
- 解釋混淆矩陣。
- 執行離線和線上模型評估 (A/B 測試)。
- 使用指標 (例如訓練模型的時間、模型品質、工程成本) 比較模型。
- 執行交叉驗證。

領域 4：機器學習實作和操作

任務說明 4.1：針對效能、可用性、可擴展性、彈性和容錯能力建立 ML 解決方案。

- 記錄和監控 AWS 環境。
 - AWS CloudTrail 和 Amazon CloudWatch
 - 建置錯誤監控解決方案。
- 部署到多個 AWS 區域和多個可用區域。
- 建立 AMI 和黃金映像。
- 建立 Docker 容器。
- 部署 Auto Scaling 群組。
- 調整資源規模 (例如執行個體、佈建 IOPS、磁碟區)。
- 執行負載平衡。
- 遵循 AWS 最佳實務。

任務說明 4.2：針對特定問題建議和實作適當的 ML 服務和功能。

- AWS 上的 ML (應用程式服務)，例如：
 - Amazon Polly
 - Amazon Lex
 - Amazon Transcribe
 - Amazon Q
- 瞭解 AWS 服務配額。
- 判斷何時建置自訂模型，以及何時使用 Amazon SageMaker 內建演算法。
- 瞭解 AWS 基礎設施 (例如執行個體類型) 和成本考量。
 - 透過 AWS Batch 使用 Spot 執行個體型訓練深度學習模型。

任務說明 4.3：將基本 **AWS** 安全實務應用至 **ML** 解決方案。

- **AWS Identity and Access Management (IAM)**
- **S3** 儲存貯體政策
- 安全群組
- **VPC**
- 加密和匿名化

任務說明 4.4：部署 **ML** 解決方案並將其操作化。

- 公開並與端點互動。
- 瞭解 **ML** 模型。
- 執行 **A/B** 測試。
- 重新訓練管道。
- 偵錯和疑難排解 **ML** 模型。
 - 偵測並減緩效能下降。
 - 監控模型效能。

附錄

考試中可能出現的技術和概念

以下清單包含考試中可能出現的技術和概念。此清單未列出所有詳情，可能會有所變更。此清單中項目的順序和位置並不代表其對應的權重或重要性：

- 擷取和收集
- 處理和 ETL
- 資料分析和視覺化
- 模型訓練
- 模型部署和推論
- 作業化 ML
- AWS 機器學習應用程式服務
- 與 ML 相關的語言 (例如 Python、Java、Scala、R、SQL)
- 筆記型電腦與整合式開發環境 (IDE)

考試範圍內的 AWS 服務和功能

下列清單包含考試範圍內的 AWS 服務和功能。此清單未列出所有詳情，可能會有所變更。AWS 產品項目會顯示在與產品主要功能一致的類別中：

分析：

- Amazon Athena
- Amazon Data Firehose
- Amazon EMR
- AWS Glue
- Amazon Kinesis
- Amazon Kinesis Data Streams
- AWS Lake Formation
- Amazon Managed Service for Apache Flink
- Amazon OpenSearch Service
- Amazon QuickSight

運算：

- AWS Batch
- Amazon EC2
- AWS Lambda

容器：

- Amazon Elastic Container Registry (Amazon ECR)
- Amazon Elastic Container Service (Amazon ECS)
- Amazon Elastic Kubernetes Service (Amazon EKS)
- AWS Fargate

資料庫：

- Amazon Redshift

物聯網：

- AWS IoT Greengrass

機器學習：

- Amazon Bedrock
- Amazon Comprehend
- AWS Deep Learning AMIs (DLAMI)
- Amazon Forecast
- Amazon Fraud Detector
- Amazon Lex
- Amazon Kendra
- Amazon Mechanical Turk
- Amazon Polly
- Amazon Q
- Amazon Rekognition
- Amazon SageMaker
- Amazon Textract
- Amazon Transcribe
- Amazon Translate

管理與控管：

- AWS CloudTrail
- Amazon CloudWatch

連網和內容交付：

- Amazon VPC

安全、身分與合規：

- AWS Identity and Access Management (IAM)

儲存：

- Amazon Elastic Block Store (Amazon EBS)
- Amazon Elastic File System (Amazon EFS)
- Amazon FSx
- Amazon S3

考試範圍以外的 **AWS** 服務和功能

下列清單包含超出考試範圍的 **AWS** 服務和功能。此清單未列出所有詳情，可能會有所變更。與考試的目標任務角色完全無關的 **AWS** 產品會從此清單中排除：

分析：

- AWS Data Pipeline

機器學習：

- AWS DeepRacer
- Amazon Machine Learning (Amazon ML)