

AWS Certified Machine Learning - Specialty

Guía de examen (MLS-C01)

Introducción

El examen AWS Certified Machine Learning - Specialty (MLS-C01) está dirigido a personas que desempeñan roles de desarrollo de inteligencia artificial o *machine learning* (IA/ML) o ciencia de datos. El examen certifica la capacidad de un candidato para diseñar, crear, implementar, optimizar, entrenar, ajustar y mantener soluciones de ML para determinados problemas empresariales mediante la nube de AWS.

En este examen, también se certifica la capacidad de un candidato para completar las siguientes tareas:

- seleccionar y justificar el enfoque de ML adecuado para un problema empresarial determinado
- identificar los servicios de AWS adecuados para implementar soluciones de ML
- diseñar e implementar soluciones de ML escalables, rentables, fiables y seguras

NOTA: El examen AWS Certified Machine Learning - Specialty está disponible actualmente en inglés, japonés, chino simplificado y coreano.

Descripción del candidato objetivo

Se espera que el candidato objetivo tenga 2 años o más de experiencia práctica en el desarrollo, la arquitectura y la ejecución de cargas de trabajo de aprendizaje profundo o ML en la nube de AWS.

Conocimientos recomendados de AWS

El candidato objetivo debe tener los siguientes conocimientos:

- la capacidad de expresar la intuición detrás de los algoritmos básicos de ML
- experiencia realizando una optimización básica de hiperparámetros
- experiencia con marcos de trabajo de aprendizaje profundo y ML
- la capacidad de seguir las prácticas recomendadas de entrenamiento con modelos
- la capacidad de seguir las prácticas recomendadas de implementación
- la capacidad de seguir las prácticas recomendadas operativas

¿Qué se considera fuera del alcance del candidato objetivo?

A continuación, se muestra una lista parcial de las tareas de trabajo relacionadas que no se espera que el candidato objetivo pueda realizar. Se considera que lo siguiente está fuera del alcance del examen:

- desarrollo de algoritmos extensos o complejos
- optimización amplia de hiperparámetros
- demostraciones y cálculos matemáticos complejos
- conocimientos avanzados en redes y diseño de redes
- conceptos avanzados de bases de datos, seguridad y DevOps
- tareas relacionadas con DevOps para Amazon EMR

Para ver una lista detallada de las herramientas y tecnologías específicas que podrían abordarse en el examen, así como listas de los servicios de AWS que se encuentran dentro y fuera del alcance, consulte el Apéndice.

Contenido del examen

Tipos de respuesta

En el examen, hay dos tipos de preguntas:

- **Opciones múltiples:** hay una respuesta correcta y tres incorrectas (distractoras)
- **Respuesta múltiple:** hay dos o más respuestas correctas entre cinco o más opciones

Seleccione una o más respuestas que completen el enunciado o respondan a la pregunta de la mejor manera. Las distractoras, o respuestas incorrectas, son opciones que podría elegir un candidato que no tenga un buen nivel de conocimientos o habilidades. Por lo general, las distractoras son respuestas verosímiles que coinciden con el área de contenido.

Las preguntas sin respuesta se califican como incorrectas. No hay penalización por adivinar. El examen incluye 50 preguntas que afectarán el puntaje.

Contenido sin puntaje

El examen incluye 15 preguntas sin puntaje que no afectan su puntuación. AWS recopila información sobre el desempeño de los candidatos en estas preguntas sin puntaje a fin de evaluarlas para su uso como preguntas con puntaje en el futuro. Estas preguntas sin puntaje no están identificadas en el examen.

Resultados del examen

El examen AWS Certified Machine Learning - Specialty (MLS-C01) es un examen que se aprueba o se desaprueba. El puntaje se obtiene según un estándar mínimo que establecen los profesionales de AWS en función de las prácticas recomendadas y las pautas del sector de la certificación.

El informe de los resultados del examen es un puntaje en la escala del 100 al 1000. La puntuación mínima para aprobar es 750. La puntuación muestra cómo le fue en el examen en general y si lo aprobó o no. Los modelos de puntajes en escala ayudan a equiparar puntuaciones de varios formatos de examen que pueden tener niveles de dificultad un poco diferentes.

El informe del puntaje puede contener una tabla de clasificación de su desempeño en cada sección. Esta información proporciona comentarios generales sobre su desempeño en el examen. En el examen, se usa un modelo de puntaje compensatorio, lo que significa que no es necesario aprobar cada sección. Solo necesita aprobar el examen general.

Cada sección del examen tiene una ponderación específica, por lo que algunas contienen más preguntas que otras. En la tabla, se presenta información general que resalta sus fortalezas y debilidades. Interprete los comentarios de cada sección con prudencia.

Descripción del contenido

En esta guía de examen, se incluyen ponderaciones, dominios de prueba y objetivos para el examen. No es una descripción completa del contenido del examen. Sin embargo, se encuentra disponible información adicional del contexto de cada uno de los objetivos a fin de ayudarlo a orientar la preparación para el examen. En la siguiente tabla, se enumeran los principales dominios de contenido y sus ponderaciones. La tabla precede a la descripción completa del contenido del examen, que incluye el contexto adicional.

Dominio	% del examen
Dominio 1: ingeniería de datos	20 %
Dominio 2: análisis exploratorio de datos	24 %
Dominio 3: modelado	36 %
Dominio 4: implementación y operaciones de <i>machine learning</i>	20 %
TOTAL	100 %

Dominio 1: ingeniería de datos

- 1.1 Crear repositorios de datos para *machine learning*
 - Identificar orígenes de datos (p. ej., el contenido y la ubicación, las fuentes principales, como los datos de los usuarios)
 - Determinar los medios de almacenamiento (p. ej., base de datos, lago de datos, S3, Elastic File System, EBS)
- 1.2 Identificar e implementar una solución de ingesta de datos
 - Estilos y tipos de trabajos de datos (carga por lotes, *streaming*)
 - Canalizaciones de ingesta de datos (cargas de trabajo de ML basadas en lotes y cargas de trabajo de ML basadas en *streaming*)
 - Kinesis
 - Kinesis Analytics
 - Kinesis Firehose
 - EMR
 - Glue
 - Programación de trabajos
- 1.3 Identificar e implementar una solución de transformación de datos
 - Transformación del tránsito de datos (ETL: Glue, EMR, AWS Batch)
 - Gestionar datos específicos de ML mediante MapReduce (Hadoop, Spark, Hive)

Dominio 2: análisis exploratorio de datos

- 2.1 Limpiar y preparar los datos para el modelado
 - Identificar y gestionar los datos faltantes, los datos dañados, las palabras vacías, etc.
 - Formatear, normalizar, aumentar y escalar datos
 - Datos etiquetados (reconocer cuándo tiene suficientes datos etiquetados e identificar las estrategias de mitigación [herramientas de etiquetado de datos (Mechanical Turk, trabajo manual)])

2.2 Realizar la ingeniería de funciones

- Identificar y extraer funciones de conjuntos de datos, incluso de orígenes de datos como texto, voz, imágenes, conjuntos de datos públicos, etc.
- Analizar o evaluar los conceptos de ingeniería de funciones (agrupación, tokenización, valores atípicos, funciones sintéticas, codificación *one-hot*, reducción de la dimensión de los datos)

2.3 Analizar y visualizar datos para *machine learning*

- Representación gráfica (diagrama de dispersión, series temporales, histograma, diagrama de caja)
- Interpretación de estadísticas descriptivas (correlación, estadísticas de resumen, valor p)
- Agrupamiento (jerárquico, diagnóstico, diagrama de codos, tamaño del clúster)

Dominio 3: modelado

3.1 Enmarcar los problemas empresariales como problemas de *machine learning*

- Determinar cuándo utilizar ML y cuándo no
- Conocer la diferencia entre el aprendizaje supervisado y el no supervisado
- Seleccionar entre clasificación, regresión, pronóstico, agrupamiento, recomendación, etc.

3.2 Seleccionar los modelos adecuados para un problema de *machine learning* determinado

- Xgboost, regresión logística, K-means, regresión lineal, árboles de decisión, bosques aleatorios, *Recurrent Neural Network* (RNN, red neuronal recurrente), *Convolutional Neural Network* (CNN, red neuronal convolucional), conjuntos, aprendizaje por transferencia
- Expresar la intuición detrás de los modelos

3.3 Entrenar modelos de *machine learning*

- Entrenar la validación de pruebas dividida y la validación cruzada
- Optimizador, descenso de gradiente, funciones de pérdida, mínimos locales, convergencia, lotes, probabilidad, etc.
- Elección de cómputos (GPU frente a CPU, distribuida frente a no distribuida, plataforma [Spark frente a no Spark])
- Actualizaciones y reentrenamiento de modelos
 - Lotes frente a tiempo real/en línea

3.4 Realizar la optimización de hiperparámetros

- Regularización
 - Abandono
 - L1/L2
- Validación cruzada
- Inicialización de modelos
- Arquitectura de redes neuronales (capas/nodos), tasa de aprendizaje, funciones de activación
- Modelos basados en árboles (número de árboles, número de niveles)
- Modelos lineales (tasa de aprendizaje)

3.5 Evaluar los modelos de *machine learning*

- Evitar el sobreajuste o el ajuste insuficiente (detectar y gestionar el sesgo y la diferencia)
- Métricas (*Area under the ROC curve* (AUC-ROC, área bajo la curva ROC), exactitud, precisión, exhaustividad, *Root Mean Square Error* [RMSE, raíz del error cuadrático medio], puntuación F1)
- Matriz de confusión

- Evaluación de modelos sin conexión y en línea, pruebas A/B
- Comparar modelos mediante métricas (tiempo de entrenamiento de un modelo, calidad del modelo, costos de ingeniería)
- Validación cruzada

Dominio 4: implementación y operaciones de machine learning

4.1 Crear soluciones de *machine learning* para el rendimiento, la disponibilidad, la escalabilidad, la resiliencia y la tolerancia a errores

- Registro y supervisión del entorno de AWS
 - CloudTrail y CloudWatch
 - Supervisión de errores de compilación
- Varias regiones, varias *availability zones* (AZ, zonas de disponibilidad)
- AMI/imagen dorada
- Contenedores de Docker
- Grupos de Auto Scaling
- Dimensionamiento correcto
 - Instancias
 - IOPS provisionadas
 - Volúmenes
- Balanceo de carga
- Prácticas recomendadas de AWS

4.2 Recomendar e implementar los servicios y funciones de *machine learning* adecuados para un problema determinado

- ML en AWS (servicios de aplicaciones)
 - Polly
 - Lex
 - Transcribe
- Límites de servicio de AWS
- Crear su propio modelo frente a algoritmos integrados de SageMaker
- Infraestructura: (*spot*, tipos de instancias), consideraciones de costos
 - Usar instancias de *spot* para entrenar modelos de aprendizaje profundo con AWS Batch

4.3 Aplicar las prácticas de seguridad básicas de AWS a las soluciones de *machine learning*

- IAM
- Políticas de *bucket* de S3
- Grupos de seguridad
- VPC
- Cifrado/anonimización

4.4 Implementar y poner en práctica soluciones de *machine learning*

- Exponer los puntos de enlace e interactuar con ellos
- Control de versiones de modelos de ML
- Pruebas A/B
- Reentrenar canalizaciones
- Depurar y solucionar problemas de ML
 - Detectar y mitigar la caída en el rendimiento
 - Supervisar el rendimiento del modelo

Apéndice

¿Qué herramientas, tecnologías y conceptos clave podrían incluirse en el examen?

A continuación, se muestra una lista parcial de las herramientas y tecnologías que podrían aparecer en el examen. Esta lista está sujeta a cambios y se proporciona para ayudarlo a comprender el alcance general de los servicios, las funciones o las tecnologías que se presentan en el examen. Las herramientas y las tecnologías generales de esta lista no aparecen en un orden particular. Los servicios de AWS se agrupan según sus funciones principales. Aunque es probable que el examen abarque algunas de estas tecnologías en mayor medida, el orden y su ubicación en esta lista no son indicios de su relevancia o importancia:

- Ingesta/recopilación
- Procesamiento/ETL
- Análisis y visualización de datos
- Entrenamiento de modelos
- Implementación e inferencia de modelos
- Conceptos operativos
- Servicios de aplicación de ML de AWS
- Lenguaje relevante para el ML (por ejemplo, Python, Java, Scala, R, SQL)
- Computadoras portátiles y entornos de desarrollo integrados (IDE)

Servicios y funciones de AWS

Análisis:

- Amazon Athena
- Amazon EMR
- Amazon Kinesis Data Analytics
- Amazon Kinesis Data Firehose
- Amazon Kinesis Data Streams
- Amazon QuickSight

Cómputo:

- AWS Batch
- Amazon EC2

Contenedores:

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

Base de datos:

- AWS Glue
- Amazon Redshift

Internet de las cosas (IoT):

- AWS IoT Greengrass

Machine learning:

- Amazon Comprehend
- AWS Deep Learning AMI (DLAMI)
- AWS DeepLens
- Amazon Forecast
- Amazon Fraud Detector
- Amazon Lex
- Amazon Polly
- Amazon Rekognition
- Amazon SageMaker
- Amazon Textract
- Amazon Transcribe
- Amazon Translate

Administración y gobernanza:

- AWS CloudTrail
- Amazon CloudWatch

Redes y entrega de contenido:

- Amazon VPC

Seguridad, identidad y conformidad:

- AWS Identity and Access Management (IAM)

Sin servidor:

- AWS Fargate
- AWS Lambda

Almacenamiento:

- Amazon Elastic File System (Amazon EFS)
- Amazon FSx
- Amazon S3

Servicios y funciones de AWS fuera de alcance

A continuación, se muestra una lista parcial de los servicios y las funciones de AWS que no se incluyen en el examen. Estos servicios y funciones no representan todas las ofertas de AWS que no se incluyen en el contenido del examen. Los servicios o las funciones que no tienen ninguna relación con los objetivos de roles de trabajo del examen se excluyen de esta lista porque se supone que son irrelevantes.

Entre los servicios y las funciones de AWS fuera de alcance se incluyen los siguientes:

- AWS Data Pipeline
- AWS DeepRacer
- Amazon Machine Learning (Amazon ML)