

AWS SageMaker

Amazon SageMaker es un servicio de aprendizaje automático (ML) totalmente gestionado. Con él SageMaker, los científicos de datos y los desarrolladores pueden crear, entrenar e implementar modelos de aprendizaje automático de forma rápida y segura en un entorno hospedado listo para la producción. Proporciona una experiencia de interfaz de usuario para ejecutar flujos de trabajo de aprendizaje automático que hace que las herramientas de aprendizaje SageMaker automático estén disponibles en varios entornos de desarrollo integrados (IDE).

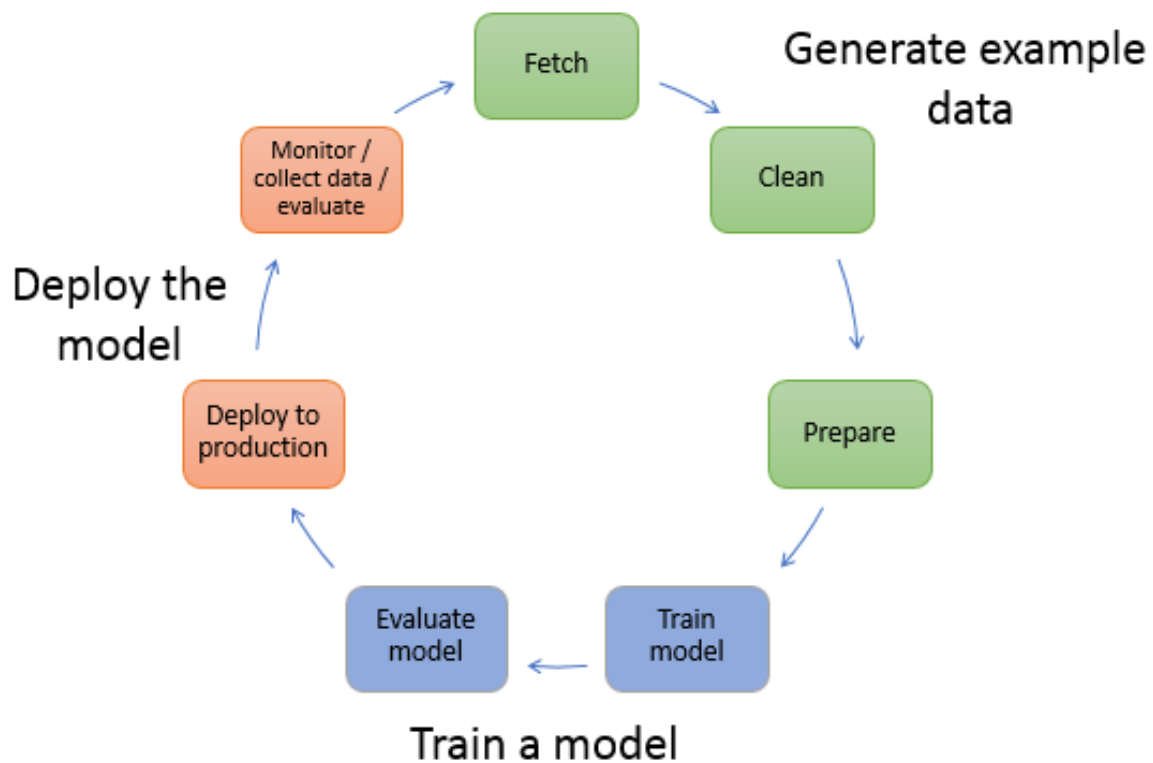
Con SageMaker ella, puede almacenar y compartir sus datos sin tener que crear y administrar sus propios servidores. Esto les da a usted o a sus organizaciones más tiempo para crear y desarrollar de forma colaborativa su flujo de trabajo de aprendizaje automático y hacerlo antes. SageMaker proporciona algoritmos de aprendizaje automático gestionados para que se ejecuten de forma eficiente con datos extremadamente grandes en un entorno distribuido. Con soporte bring-your-own-algorithms y marcos integrados, SageMaker ofrece opciones de formación distribuidas y flexibles que se ajustan a sus flujos de trabajo específicos. En unos pocos pasos, puede implementar un modelo en un entorno seguro y escalable desde la SageMaker consola.

Descripción general del aprendizaje automático con Amazon SageMaker

En esta sección se describe un flujo de trabajo típico de aprendizaje automático (ML) y se resume cómo realizar esas tareas con Amazon SageMaker.

En el aprendizaje automático, enseñas a un ordenador a hacer predicciones o inferencias. En primer lugar, utilice el algoritmo o los datos de ejemplo para realizar la capacitación de un modelo. A continuación, integre el modelo en la aplicación para generar inferencias en tiempo real y a escala.

El siguiente diagrama ilustra el flujo de trabajo típico para crear un modelo de aprendizaje automático. Incluye tres etapas en un flujo circular que trataremos con más detalle a continuación: generar datos de ejemplo, entrenar un modelo e implementar el modelo.



El diagrama ilustra cómo realizar las siguientes actividades en la mayoría de los escenarios típicos:

Genere datos de ejemplo: para entrenar un modelo, necesita datos de ejemplo. El tipo de datos que necesita depende del problema empresarial que desee que el modelo resuelva (las inferencias que desea que genere el modelo). Por ejemplo, supongamos que desea crear un modelo para predecir un número a partir de una imagen de entrada de un dígito manuscrito. Para realizar la capacitación de ese modelo, necesita ejemplos de imágenes de números escritos a mano.

Los científicos de datos suelen dedicar tiempo a explorar y preprocesar datos de ejemplo antes de usarlos para el entrenamiento de modelos. Para preprocesar datos, normalmente realiza la siguiente operación:

Obtenga los datos: es posible que tenga repositorios de datos de ejemplo internos o que utilice conjuntos de datos que estén disponibles públicamente. Normalmente, se extraen el conjunto o los conjuntos de datos en un repositorio único.

Limpie los datos: para mejorar el entrenamiento del modelo, inspeccione los datos y límpielos, según sea necesario. Por ejemplo, si los datos tienen un country name atributo con valores United StatesUS, puede editarlos para que sean coherentes.

Prepare o transforme los datos: para mejorar el rendimiento, puede realizar transformaciones de datos adicionales. Por ejemplo, puede elegir combinar atributos. Si su modelo predice las condiciones que requieren el deshielo de una aeronave, en lugar de utilizar los atributos de temperatura y humedad por separado, puede combinar esos atributos en un nuevo atributo para obtener un modelo mejor.

En SageMaker, puede preprocesar datos de ejemplo mediante SageMaker API con el SDK de SageMaker Python en un entorno de desarrollo integrado (IDE). Con el SDK para Python (Boto3) puedes recuperar, explorar y preparar tus datos para el entrenamiento de modelos. Para obtener información sobre la preparación, el procesamiento y la transformación de los datos [Preparación de los datos](#), [Procesamiento de datos](#) consulte [y. Crea, almacena y comparte funciones con Amazon SageMaker Feature Store](#)

Entrenar un modelo: el entrenamiento del modelo incluye tanto la capacitación como la evaluación del modelo, de la siguiente manera:

Entrenamiento del modelo: para entrenar un modelo, se necesita un algoritmo o un modelo base previamente entrenado. El algoritmo que elija dependerá de un número de factores. Para una solución integrada, puede utilizar uno de los algoritmos que se SageMaker proporcionan. Para obtener una lista de los algoritmos proporcionados por SageMaker y las consideraciones relacionadas, consulte [Usa algoritmos SageMaker integrados de Amazon o modelos previamente entrenados](#). Para obtener una solución de entrenamiento basada en la interfaz de usuario que proporciona algoritmos y modelos, consulte [SageMaker JumpStart](#).

También necesita computar los recursos para la capacitación. Según el tamaño del conjunto de datos de capacitación y la rapidez con la que necesite los resultados, puede utilizar recursos desde una instancia de propósito general única a un clúster distribuido de instancias de GPU. Para obtener más información, consulte [Entrena a un modelo con Amazon SageMaker](#).

Evaluación del modelo: después de entrenar el modelo, se evalúa para determinar si la precisión de las inferencias es aceptable. Para entrenar y evaluar su modelo, puede usar el SDK de SageMaker Python para enviar solicitudes de inferencias al modelo a través de uno de los IDE disponibles. Para obtener más información sobre la evaluación del modelo, consulte [Supervisión de la calidad de los datos y los modelos](#).

Implemente el modelo: tradicionalmente, se rediseña un modelo antes de integrarlo con la aplicación e implementarlo. Con los servicios de SageMaker alojamiento, puede implementar el modelo de forma independiente, lo que lo desvincula del código de la aplicación. Para obtener más información, consulte Implementar modelos para inferencia.

El aprendizaje automático es un ciclo continuo. Tras implementar un modelo, se supervisan las inferencias, se recopilan más datos de alta calidad y se evalúa el modelo para identificar desviaciones. A continuación, aumenta la precisión de las inferencias actualizando los datos de entrenamiento para incluir los datos de alta calidad recién recopilados. A medida que haya más datos de ejemplo disponibles, seguirás reentrenando tu modelo para aumentar la precisión.

Algoritmos SageMaker integrados de Amazon o modelos previamente entrenados

https://docs.aws.amazon.com/es_es/sagemaker/latest/dg/algos.html