

¿En qué tipo de problemas es adecuado utilizar aprendizaje automático?

- Clasificación
- Regresión
- Agrupamiento
- Recomendación

¿Cómo funciona un problema de clasificación?

Se necesita clasificar elementos en categorías discretas y predefinidas

Extra: clasificar correos electrónicos como spam o no spam, clasificar imágenes según su contenido o clasificar clientes en diferentes grupos demográficos.

¿Cómo funciona un problema de regresión?

Se necesita predecir un valor basado en datos históricos

Extra: predecir el precio de una casa en función de sus características, predecir el tiempo que tardará un cliente en pagar una factura o predecir el número de ventas de un producto en función de la publicidad.

¿Cómo funciona un problema de agrupamiento?

Se necesita agrupar elementos similares en grupos

Extra: agrupar clientes en función de sus preferencias, agrupar documentos en función de su contenido o agrupar imágenes en función de su contenido.

¿Cómo funciona un problema de recomendación?

Se necesita recomendar elementos a los usuarios en función de sus preferencias

Extra: recomendar películas a los usuarios en función de sus calificaciones, recomendar productos a los clientes en función de sus compras anteriores o recomendar música a los usuarios en función de sus preferencias.

¿Qué representa el resultado de la función de costo o pérdida?

Que tan mala es la predicción

¿Qué toma como input la función de costo o pérdida?

El valor predicho y el valor real

¿Cómo se utiliza la función de pérdida para entrenar un modelo?

Se busca minimizar la función de pérdida

¿Qué función se utiliza para calcular la pérdida en un problema de regresión?

Error cuadrático medio

¿Cuál es la característica de la función de pérdida? (Frente a la métrica de rendimiento)

Es fácil de minimizar y no presenta mucho overhead computacional

¿Por qué no se usa la métrica de rendimiento para entrenar un modelo?

No siempre es fácil de minimizar computacionalmente.

¿Por qué no se usa la función de pérdida para evaluar un modelo?

No siempre es fácil de interpretar sus resultados para los humanos

¿Cuál es la característica de la métrica de rendimiento? (Frente a la función de pérdida)

Es fácil de interpretar para los humanos

¿Qué función se utiliza como métrica de rendimiento en un problema de regresión?

la raíz del error cuadrático medio

¿Cuál es la fórmula del error cuadrático medio?

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

¿Cuál es la fórmula de la raíz del error cuadrático medio?

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

¿Que estrategias existen para entrenar un modelo utilizando features o características no numéricas?

- One-hot encoding
- Label encoding
- Ordinal encoding

¿Cómo funciona el one-hot encoding?

Se crea una columna por cada valor posible de la característica y se asigna un 1 a la columna correspondiente al valor de la característica y un 0 a las demás columnas

¿Cuando se utiliza el one-hot encoding?

Cuando no existe un orden entre los valores de la característica

¿Que desventaja tiene el one-hot encoding?

Aumenta la dimensionalidad del conjunto de datos

¿Cómo funciona el ordinal encoding?

Se asigna un número a cada valor de la característica en función de su orden

¿Cuando se utiliza el ordinal encoding?

Cuando existe un orden entre los valores de la característica

Extra: si tenemos una característica "Tamaño" con categorías "Pequeño", "Mediano" y "Grande", podemos asignar los valores 0, 1 y 2 respectivamente

¿Cómo funciona el label encoding?

Se asigna un número a cada valor de la característica sin considerar el orden entre las categorías

Extra: El número es arbitrario

¿Cuando se utiliza el label encoding?

Cuando no existe un orden entre los valores de la característica y se busca reducir la dimensionalidad del conjunto de datos

¿Por qué no se puede usar un método analítico minimizando la derivada de la función para minimizar la función de costo?

Porque la función de costo puede ser muy compleja y derivarla puede ser muy costoso computacionalmente o incluso imposible

¿Por qué se usa un método iterativo para minimizar la función de costo?

- Porque es computacionalmente mucho más eficiente que el método analítico
- Su resultado es lo suficientemente bueno (Aunque no sea exacto)

¿Por qué el algoritmo SGD se llama “estocástico”?

Porque utiliza un subconjunto aleatorio de los datos de entrenamiento en cada iteración para calcular el gradiente

¿Qué sucede si tenemos un learning rate muy grande?

El algoritmo puede no converger adecuadamente y oscilar alrededor del mínimo global o diverger

¿Qué sucede si tenemos un learning rate muy pequeño?

Convergencia muy lenta, requiriendo un mayor número de iteraciones

¿Qué determina el valor del learning rate?

la magnitud de los ajustes realizados en los parámetros del modelo en cada iteración del algoritmo

Si una red tiene muy buen rendimiento durante el entrenamiento, pero muy bajo rendimiento en un conjunto de datos de evaluación ¿Qué tipo de problema es?

Overfitting

¿Cómo se soluciona el overfitting?

- Eliminando features
- Agregando regularización

¿Cómo se soluciona el underfitting?

- Agregando features
- Aumentando la complejidad del modelo (Usando un modelo no lineal)

¿Por qué inicializamos los pesos de un modelo en valores aleatorios? (4)

- Ayudar al modelo a escapar de mínimos locales
- Encontrar mejores soluciones durante la optimización
- Explorar diferentes regiones del espacio de pesos
- Evitar gradient vanishing o gradient exploding

¿Qué pasaría si inicializamos los pesos de un modelo todos en cero?

El modelo no sería capaz de aprender porque el gradiente siempre sería nulo

¿Qué pasaría si inicializamos los pesos de un modelo todos en uno?

Los gradientes serían iguales y el modelo no sería capaz de aprender

¿Por qué el dropout puede ser considerado una técnica de regularización?

Porque reduce la complejidad del modelo

¿Por qué está recomendada la estandarización de los valores de entrada de un modelo de machine learning?

Porque ayuda a que el algoritmo de optimización converja más rápido