

# Fundamentos de la ciencia de datos para el aprendizaje automático por Microsoft

## Módulo 1:

### Introducción al aprendizaje automático:

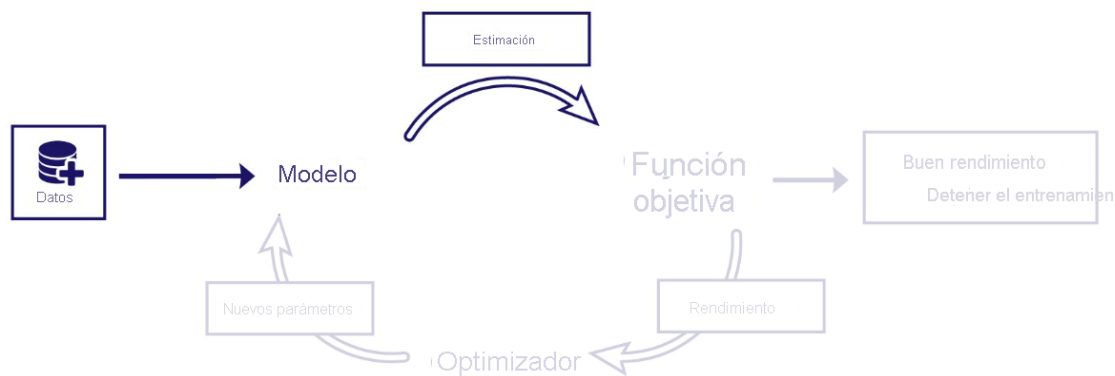
- Explorar la diferencia entre el aprendizaje automático y el software tradicional
  - Crear y entrenar un modelo de aprendizaje automático
  - Cargar un modelo y usarlo con datos nuevos
1. Selección de un modelo:
    - a. El primer paso del ML es seleccionar el tipo de modelo que se quiere usar.
  2. Detección de los parámetros durante el entrenamiento:
    - a. Se define una estimación inicial de los valores de parámetros y, después, dichos parámetros se ajustan durante el proceso de aprendizaje automatizado denominado entrenamiento

### Input and Output

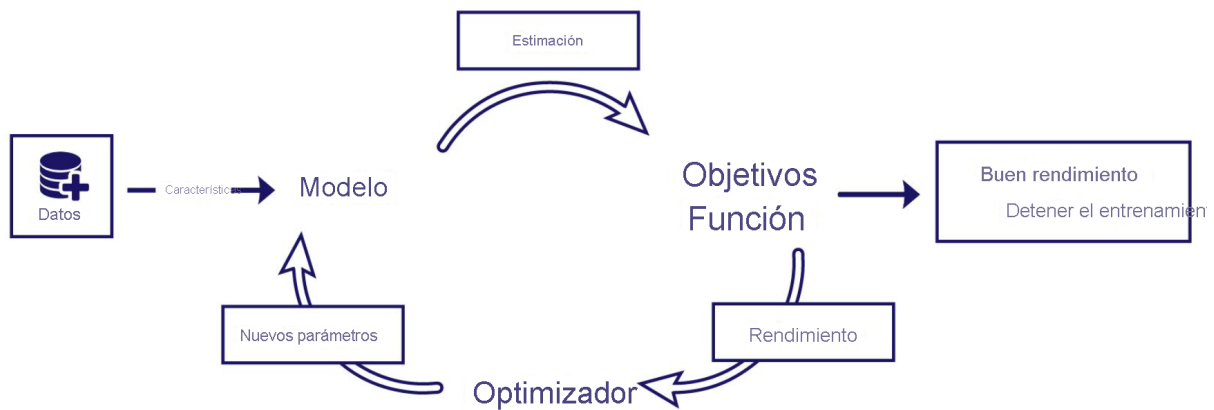
Se entrenan con datos y con dos fragmentos de código: la función objetivo (también denominado función costo) y el optimizador

### Diferencias entre entrenar y usar un modelo

Usar un modelo significa proporcionarle entradas y recibir predicciones o estimaciones. Esto se utiliza mediante el entrenamiento con datos reales. Esto tarda segundos o minutos.



Por lo contrario, entrenar un modelo es el proceso de mejorar el funcionamiento de dicho modelo, esto requiere el uso del modelo, la función objetivo y el optimizador en un bucle específico. Esto puede tardar unos minutos o hasta inclusive días.



## Módulo 2:

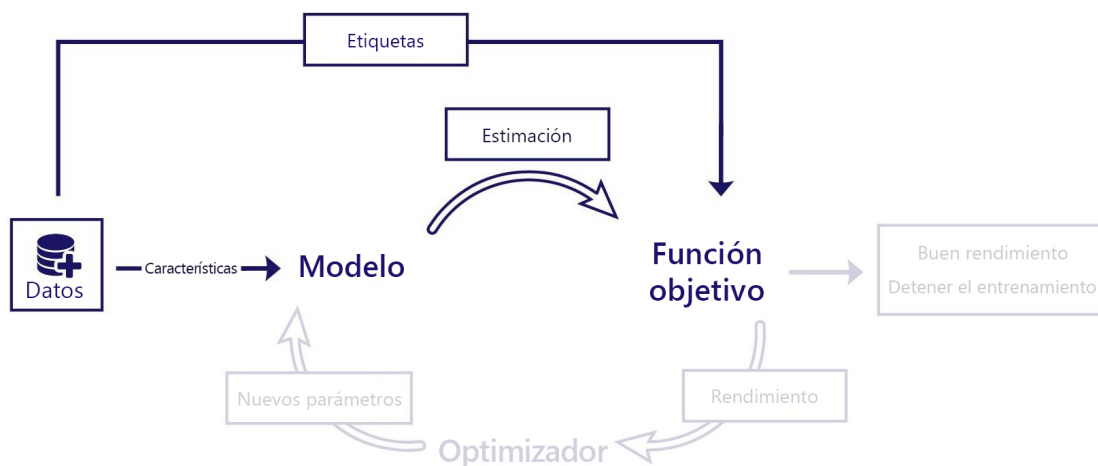
### Creación de modelos de ML clásicos con aprendizaje supervisado

- Definir el aprendizaje supervisado y sin supervisión.
- Explorar cómo afectan las funciones de costo al proceso de aprendizaje.
- Descubrir cómo se optimizan los modelos mediante el descenso de gradiente.
- Experimentar con las velocidades de aprendizaje y ver cómo pueden afectar al entrenamiento.

#### Definición del aprendizaje supervisado

Se requiere de lo siguiente en la función objetivo para determinar un modelo como aprendizaje supervisado:

- Características que se proporcionan como entradas
- Etiquetas, que son las respuestas correctas que queremos que el modelo pueda generar



Ejemplo: consideremos que vamos a predecir qué temperatura se alcanzará el 31 de enero de un año determinado.

Para esta predicción, necesitamos dos datos importantes:

- Características: fecha
- Etiqueta: temperatura diaria (por ejemplo, de registros históricos)

Después del entrenamiento la etiqueta no se llega utilizar solo durante

Reducción de los errores del modelo con funciones de costo

Error, costo y pérdida

Hace referencia al número de errores que un modelo comete al predecir una o varias etiquetas

La reducción del costo es nuestro objetivo

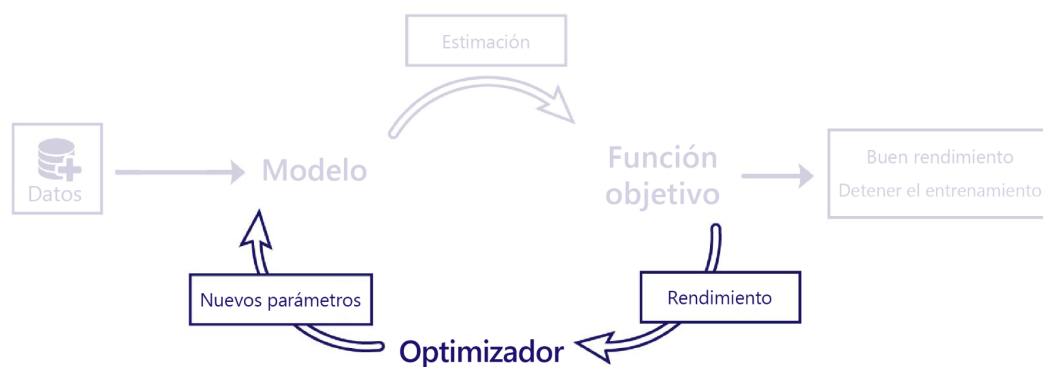
Necesitamos entrenar un modelo hasta llegar a un costo 0(sin ningún error), en lo cual es algo difícil de concluir por eso se determina el modelo a partir de la tasa de errores que tendrá mediante el modelado

¿Qué es una función de costo?

Es un pequeño código donde determina a partir de la predicción del modelo de aprendizaje, podemos tomar una decisión a partir de la función costo, si no tenemos buenos resultados podemos pasarle la información al optimizador

Optimización de modelos mediante el descenso de gradiente

El rol del optimizador es modificar el modelo de forma que mejore el rendimiento. Para ello, altera las salidas y el costo del modelo y sugiere nuevos parámetros para dicho modelo



Descenso de gradiente

El algoritmo de optimización más común es el descenso de gradiente

Calcula como los cambios realizados en cada parámetro modifican el coste mediante el uso del cálculo.

Este algoritmo es sencillo pero eficaz. Sin embargo, no garantiza encontrar unos parámetros del modelo óptimos que minimizan el costo. Las dos principales fuentes de error son los mínimos locales y la inestabilidad.

#### ─Mínimos locales

Es propenso a buscar los mínimos locales: estimaciones de parámetros que no son la mejor solución, pero cuyo gradiente es cero

Inestabilidad:

Esta inestabilidad suele ocurrir cuando el tamaño del paso o la velocidad de aprendizaje (cuánto se ajusta  $c$ /parámetro en  $c$ /iteración) es demasiado grande

### Módulo 3:

#### Introducción a los datos para el aprendizaje automático

- Visualización de conjuntos de datos grandes con análisis exploratorio de los datos(EDA)
- Limpiar los errores de un conjunto de datos
- Predecir valores desconocidos con datos numéricos y categóricos

#### ─Datos correcto e incorrectos y datos que faltan

Necesitamos muestras de datos voluminosas y representativas que:

- Tengas cero errores
- Contengan toda la información clave

#### ─“Representativas” ¿Qué significan?

Tiene que tener el ámbito estadístico dos conceptos claves, la población, y muestras para decidir si los recursos de datos disponibles realmente ayudan al análisis y estudio.

#### ─Error de datos

Suficientes datos erróneos pueden sesgar un modelo y llega a realizar predicciones erróneas

Se pueden agrupar en dos categorías:

- Errores de medición: hace referencia a datos de mediciones de baja calidad en la fase de recopilación de los datos; Estos datos suelen ser sutiles y difíciles o imposibles de eliminar
- Errores de entrada: hace referencia a datos recopilados con precisión que después se ponen de forma incorrecta o inexacta en una hoja de cálculo o recurso administrativo de datos. Estos son más fáciles de detectar

#### ─Para datos incompletos, podemos

- Elegir un modelo que funcione con datos incompletos, o bien;
- Quitar las muestras(filas) que tienen datos incompletos y trabajar con los datos que pueden.
- Agregar valores artificialmente como sustitutos razonables de los datos que faltan.

### -Datos continuos, ordinales y categóricos

Conocer lo que representa los datos, nos permite a elegir de manera adecuada el siguiente modelo que pueden adaptarse, también nos permite a organizar de manera específica y útil

-Datos continuos: Hace referencia a información numérica que puede aumentar o disminuir en cualquier cantidad. Ejemplo; puede agregar 1 milímetro a 1 metro y calcular una suma como 1,001 metros.

-Datos categóricos: Hace referencia a los datos que no están en un aspecto continuo.

Ejemplo: las personas del Titanic se clasifican como “tripulación” o “pasajeros”.

Los datos categóricos no se pueden almacenar como números de una forma obvia.

-Datos ordinales: Hace referencia a datos categóricos que tienen un orden definido y, por tanto, se pueden almacenar como valores numéricos; por ejemplo, podemos determinar por “grande”, “mediano”, “pequeño”; lo transformamos a datos numéricos pero “manzanas”, “peras”, “banana” son datos categóricos pero no lo podemos transformar en datos numéricos

Los datos ordinales también hacen referencia a números que se pueden incrementar o reducir, pero solo en cantidades establecidas. Por ejemplo, el número que suben a un barco es un número entero y incrementa, pero no se puede establecer un embarcó a medias

-ID: Dato para determinar y poder navegar mediante ellos a determinar más rápido cualquier problema que se nos presente

### -Tipo de datos:

- Int: números naturales, por ejemplo 2
- Float: números con posiciones decimales: por ejemplo, 1,2
- String: letras y palabras
- Booleanos : True o False
- Null: no son datos, sino ausencias de las mismas

### -Tipos de datos derivados

Los equipos de datos pueden almacenar fechas, imágenes y modelos 3D, entre otros a esto se hace referencia los datos derivados

Es muy común transformar estos datos en datos más simples: por ejemplo, tenemos una fecha “1 de enero del 2000” lo transformamos a un número flotante 1012000 de esta manera es más fácil de aprovechar este dato

### -¿Demasiadas opciones?

- Para trabajar con datos continuos, los números float son la mejor opción
- Los datos ordinales se suelen codificar con valores int
- Normalmente, los datos categóricos que solo implican dos categorías pueden codificarse con datos booleanos o int. Trabajar con tres categorías o más puede ser algo complicado

### -Vectores One-Hot

Vamos a codificar los datos anteriores.

Los datos categóricos no son números

Estos datos no funcionan con números del mismo modo que otros tipos de datos funcionan con números, ejemplo: en el Titanic, un precio de billete de 30\$ es más caro que el de un billete de 12\$

Por lo contrario, los datos categóricos no tienen un orden lógico. Tendremos problemas si intentamos modificar como características categóricas que tenga más de dos clases

Ejemplo: puerto de embarque tiene tres valores: C(Cherbourg), Q(Queenstown), S(Southampton). No podemos reemplazar estos símbolos por números; si lo hacemos implicaría una comparación

La codificación one-hot permite codificar estos datos categóricamente no lógicos para poder obtener mayor información

## Módulo 4:

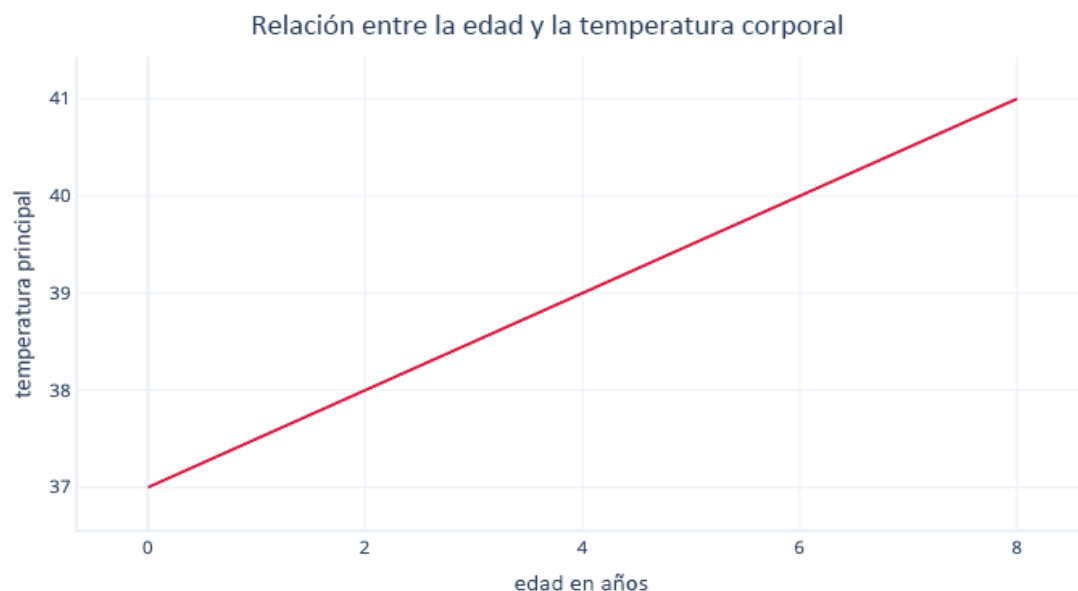
### Aprendizaje y conocimiento de los modelos de regresión en el aprendizaje automático

- Comprender el funcionamiento de la regresión
- Trabajar con algoritmos nuevos: regresión lineal, regresión lineal múltiple y regresión polinómica.
- Comprender los puntos fuertes y las limitaciones de los modelos de regresión.
- Visualizar las funciones de error y costo en la regresión lineal.
- Comprender las métricas de evaluación básicas para la regresión

La regresión es una técnica de análisis de datos sencilla, común y muy útil, a menudo denominada coloquialmente “ajustar una línea”. En su forma más simple, la regresión ajusta una línea recta entre una variable (característica) y otra(etiqueta).

#### -Regresión lineal simple

Modela una relación lineal entre una sola característica y una etiqueta normalmente continua, lo que permite que las características prediga la etiqueta, tiene un aspecto similar al siguiente:



Los parámetros: una intersección(c), que indica el valor de la etiqueta cuando la característica se establece en 0, y una pendiente (m), que indica cuánto aumentará la etiqueta para cada incremento de un punto de la característica  
Matemáticamente se vería tal que así:

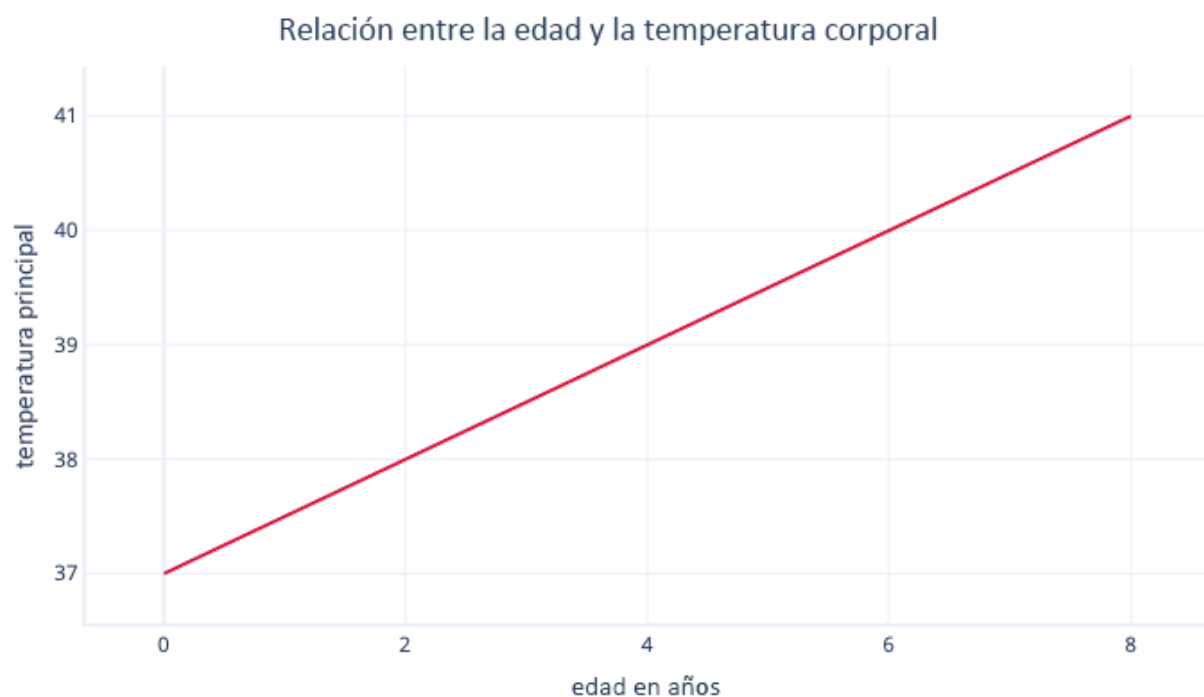
$$y = mx + c$$

y: etiqueta

x: característica

Por ejemplo, en nuestro escenario, si intentamos predecir qué pacientes tendrán temperatura corporal elevada en función a su edad, tendríamos el siguiente modelo:  
temperatura = m \* edad + c

Debe buscar los valores de m y c durante el procedimiento de ajuste, si suponemos que m=0.5 y c=37, es posible que nos quede el gráfico de la siguiente manera

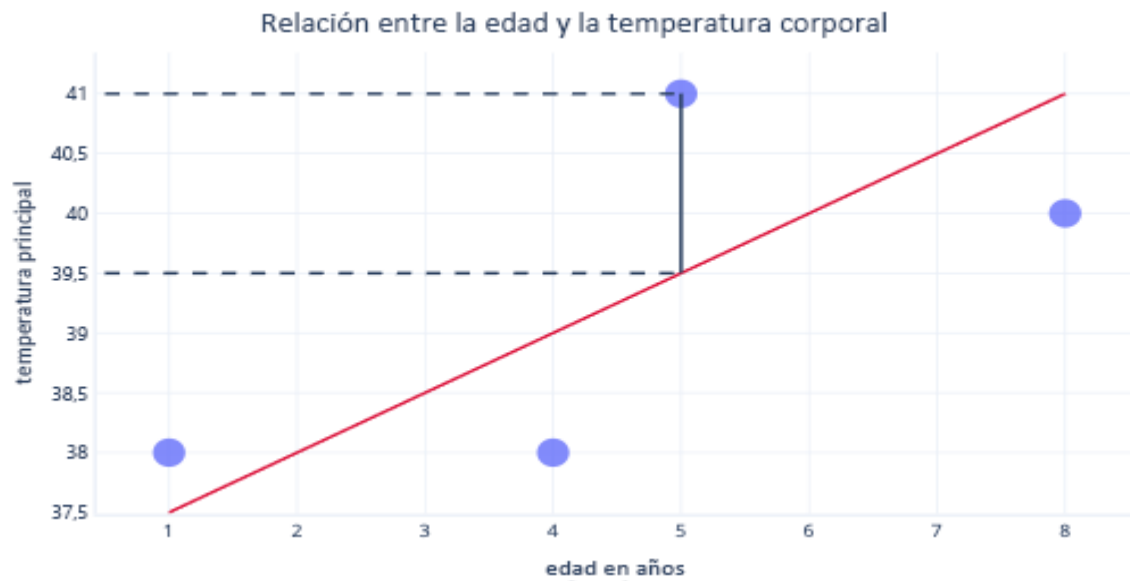


Podemos observar que c/año de edad está asociado con un aumento de la temperatura corporal de 0,5°C, con punto inicial de 37°C

#### Ajuste de la regresión lineal

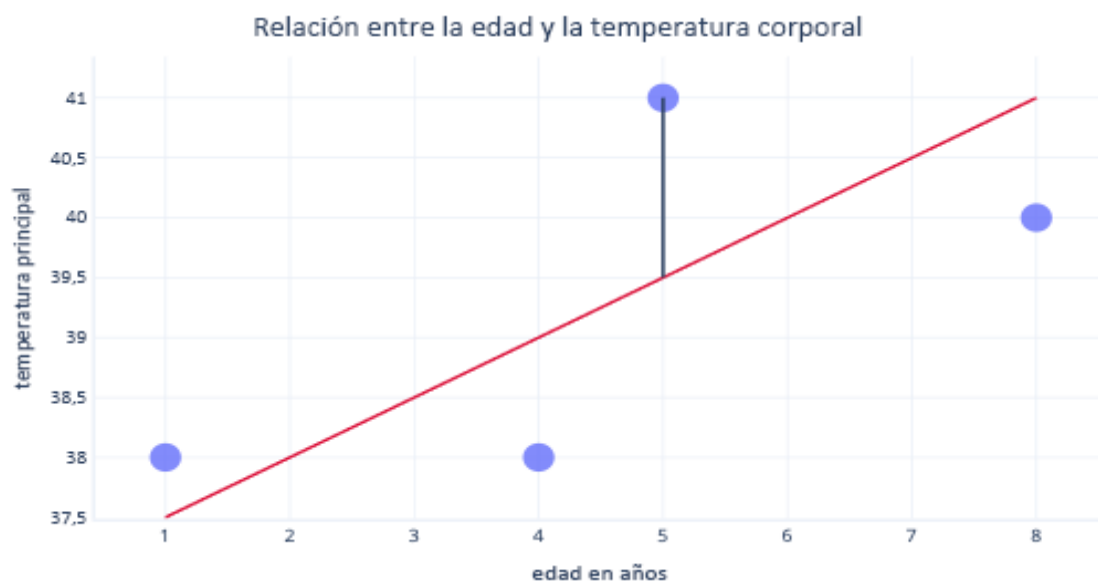
Normalmente, usamos bibliotecas existentes para ajustarnos a los modelos de regresión.

La regresión normalmente pretende encontrar la línea que genera el error mínimo, donde se refiere a la diferencia entre el valor punto de datos real y el valor previsto. Por ejemplo en la imagen siguiente, la línea negra indica el error entre la predicción, la línea roja y un valor real: el punto



AI

mirar estos dos puntos en un eje Y, podemos ver que la predicción era 39.5, pero el valor real era 41



PLT,

el modelo estaba equivocado en 1,5

Normalmente, ajustamos un modelo minimizando la suma residual de cuadrados.

La función costo se calcula de la siguiente manera:

- Calcule la diferencia entre los valores reales y los previstos (como los anteriores) para cada punto de datos.
- Eleve estos valores al cuadrado
- Sume (o halle el promedio) de estos valores al cuadrado

▬Puntos fuertes de la regresión

Las técnicas dispone de múltiples ventajas a comparación de otros modelos más complejos

▬Fácil de extrapolar



Las regresiones hacen que sea más fácil extrapolar: realizar predicciones para valores fuera del intervalo de nuestro conjunto de datos

Normalmente, se garantiza un ajuste óptimo

La mayoría de los modelos de aprendizaje automático usan el trazo descendente del gradiente para ajustar el modelo

Regresión lineal múltiple y  $R^2$

La regresión lineal múltiple modela la relación entre varias características y una sola variable

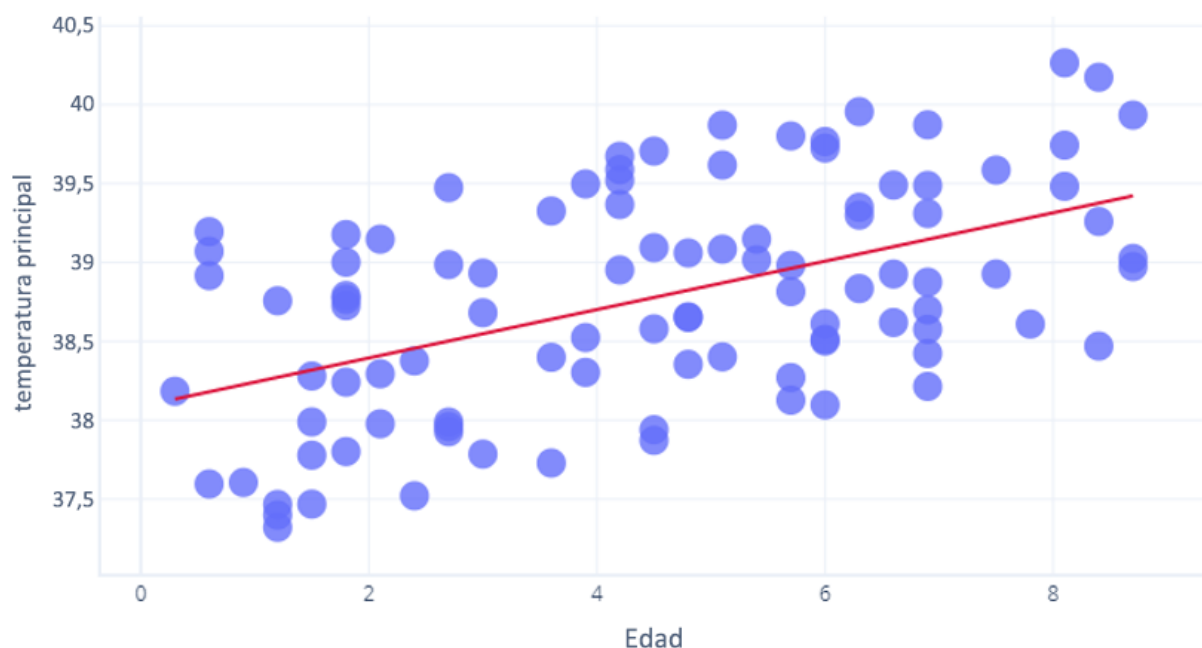
La regresión lineal múltiple tiene suposiciones

El hecho de que el modelo espere que las características sean independientes se denomina suposición de modelo. Cuando las suposiciones del modelo no son ciertas, las predicciones del modelo pueden ser erróneas

Bondad de ajuste:  $R^2$

Es un valor entre 0 y 1 que nos indica cómo se ajusta un modelo de regresión lineal a los datos

En este caso la relación entre *age\_core* y *core\_temperature*, significaba que si tiene relación sería 1 y si no 0 gráficamente sería lo siguiente:



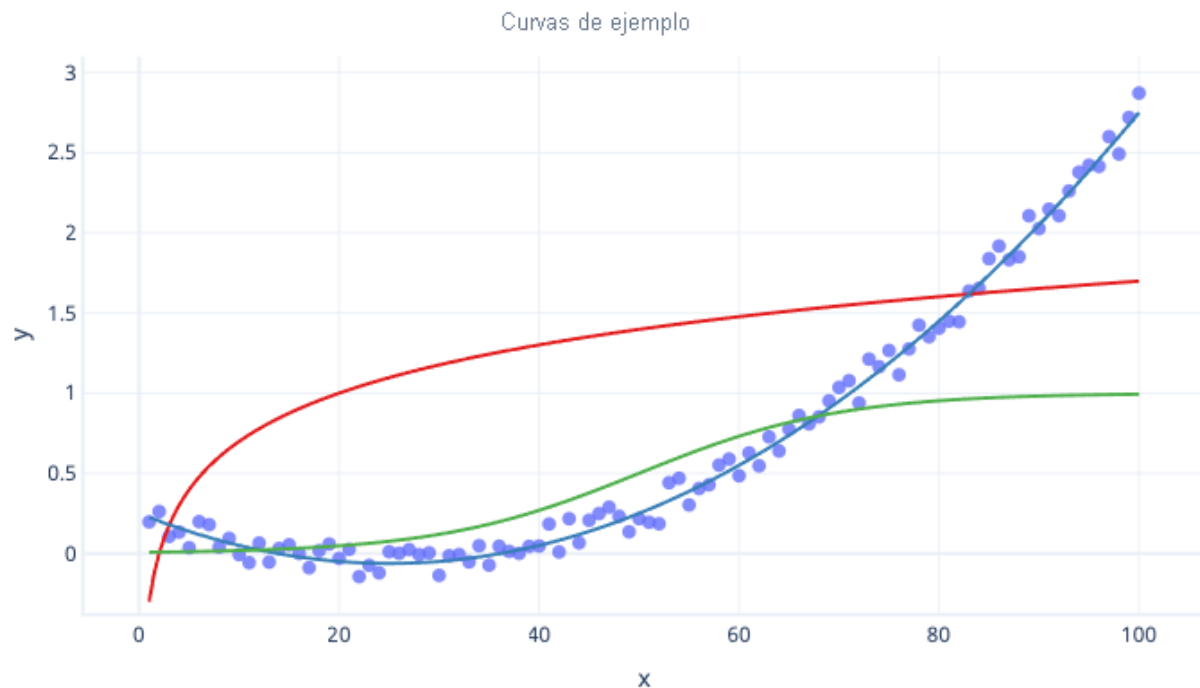
Esta herramienta tiene grandes limitaciones

- Debido a cómo se calcula  $R$  cuantas más muestras tenemos, mayor será el  $R$ . Esto hace pensar que un modelo es mejor que el otro (idéntico)
- Los valores de  $R$  no nos dicen cómo se verán con datos nuevos ya que solo utiliza datos previos
- Los valores de  $R$  no nos dicen la dirección de relación, no nos indica si está creciendo o decreciendo la pendiente

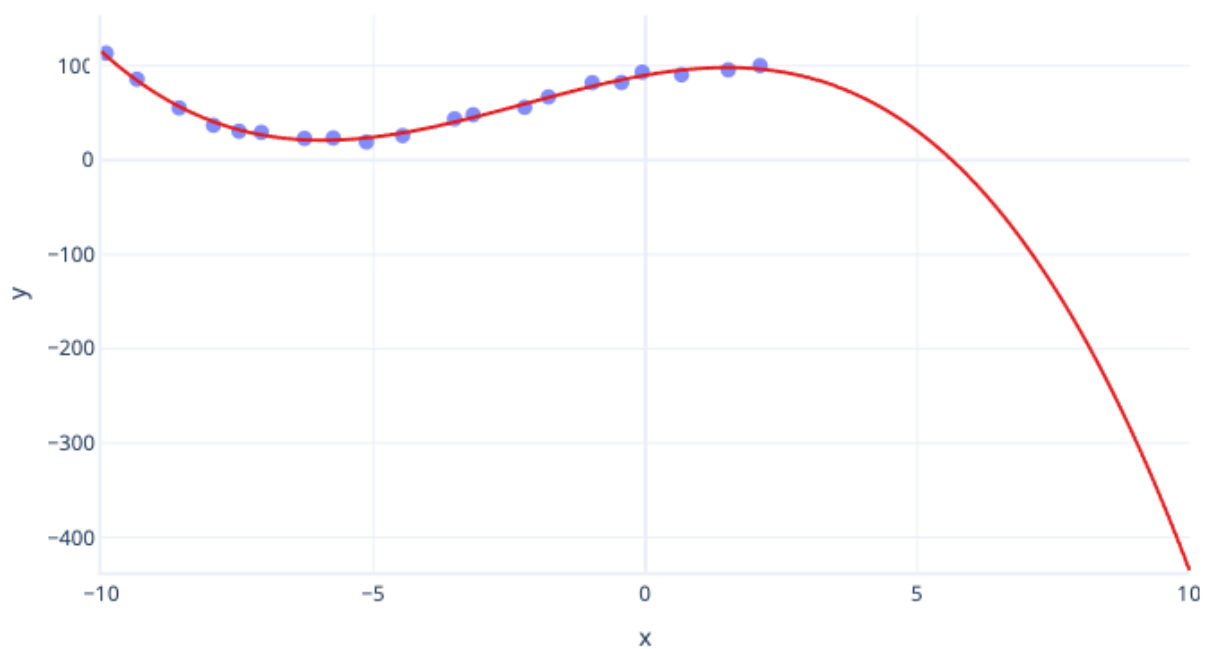
### Regresión lineal polinómica

Modela las relaciones como tipo determinado de curva

La ventaja es que se puede usar para observar todo tipo de relaciones. Por ejemplo, se puede usar para relaciones que son negativas dentro de un determinado intervalo de valores de características, pero positivas en otras



La desventaja principal de las curvas polinómicas es que se extrapolan de manera errónea. En otras palabras, si intentamos predecir valores mayores o menores que nuestros datos de entrenamiento, los polinomios pueden predecir valores extremos muy pocos realistas. Otra desventaja es que las curvas polinómicas son fáciles de sobreajuste



## Módulo 5:

### Creación y descripción de modelos de clasificación en el aprendizaje automático

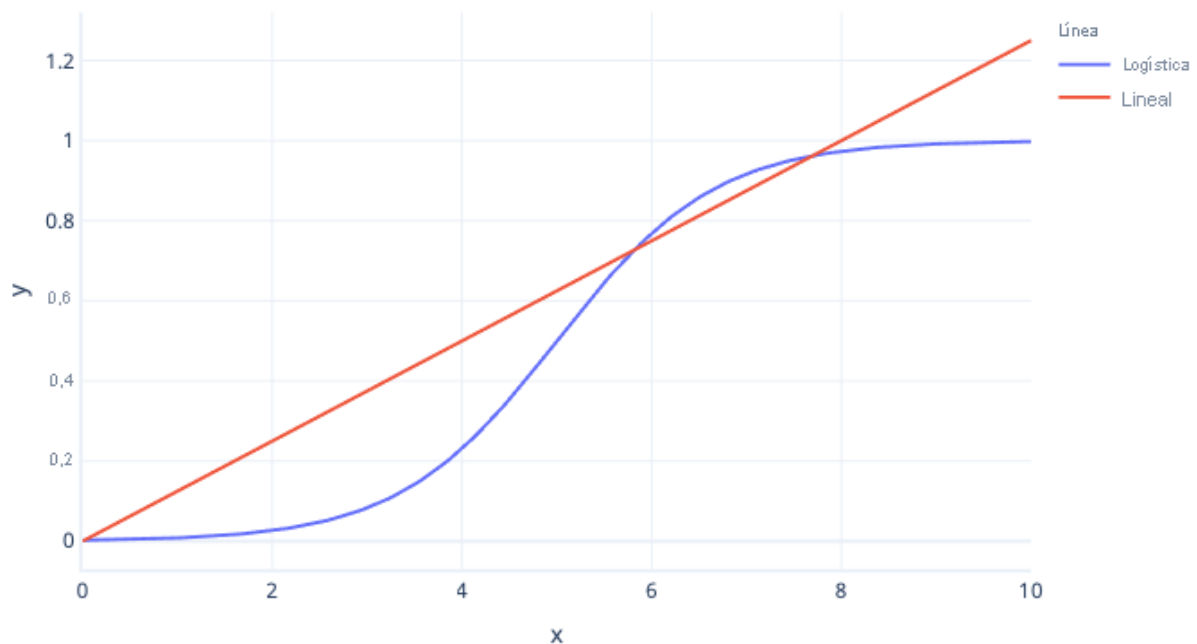
- Diferencias de la clasificación de la regresión clásica
- Modelos de compilación que pueden realizar tareas de clasificación
- Explorar cómo evaluar y mejorar los modelos de clasificación

¿Qué son los modelos de clasificación?

Los modelos de clasificación se usan para tomar decisiones o asignar elementos a categorías. A diferencia de los modelos de regresión, que emiten números continuos, como alturas o pesos, los modelos de clasificación emiten booleanos (True or False) o decisiones categóricas, como “apple”, “banana”, “cherry”

Regresión logística

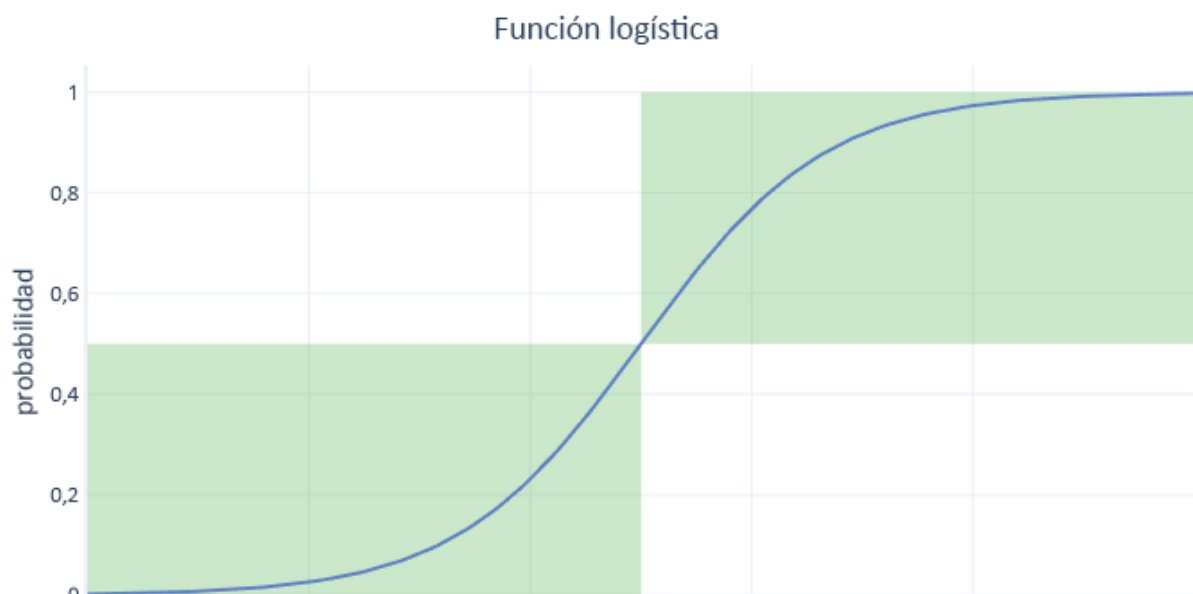
Uno de los mejores modelos de clasificación para aprender es la regresión logística; Funciona de forma similar a la regresión lineal. La diferencia entre esta y la regresión lineal es la forma de la curva. Podremos apreciarlo en el siguiente gráfico:



La regresión logística es mejor para estimar los resultados booleanos que la regresión lineal, porque la curva logística siempre retorna un valor entre 0(False) y 1(True). Cualquier valor que esté dentro de estos dos valores se considera una probabilidad.

#### Conversión de salida en categorías

Dado que la regresión logística proporciona estas probabilidades, en lugar de valores true o false simples, es necesario realizar pasos adicionales para convertir el resultado en una categoría. La manera más sencilla de hacerlo es aplicar un umbral. Por ejemplo, en el gráfico siguiente, el umbral se establece en 0,5. Este umbral significa que cualquier valor de Y por debajo de 0,5 es false y cualquier valor por encima de 0,5 es true



### ¿Evaluación de un modelo de clasificación

Los modelos de clasificación necesitan evaluación, al igual que los modelos de regresión, aunque la forma que se lleva esta evaluación es mucha más compleja.

### ¿Recordatorio sobre el costo

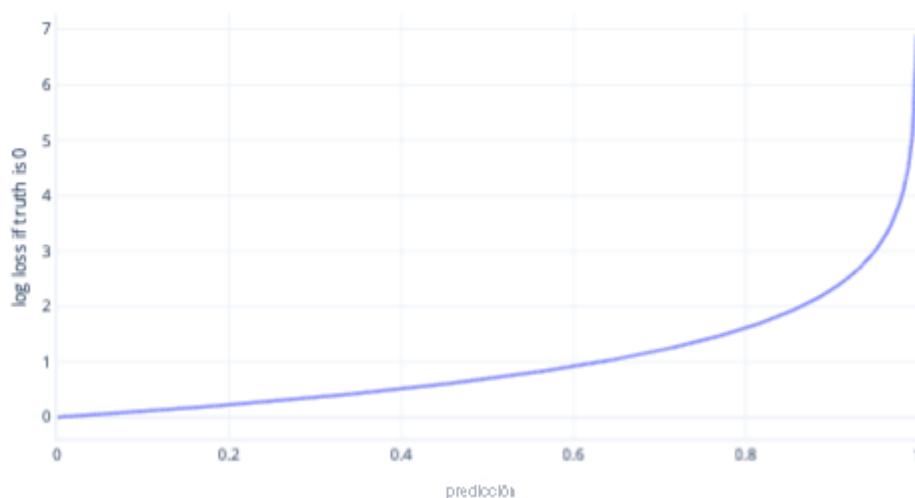
Recordemos que durante el entrenamiento, calculamos el bajo rendimiento de un modelo y lo denominamos costos o pérdidas. Podríamos usar MSE (Error Cuadrático Medio)

### ¿Funciones de costo para la clasificación

Los modelos se juzgan a partir de su probabilidad de la salida

### ¿Pérdida de costo para la clasificación

Esta función es la más popular para la clasificación simple

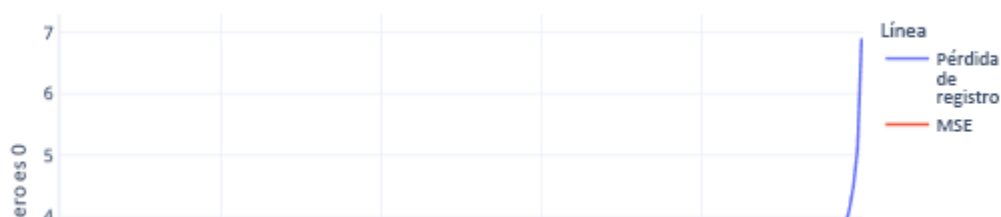


En el eje X hay posibilidades salidas del modelo (probabilidades de 0 a 1) y el eje Y indica el costo. Si un modelo tiene una confianza alta de que la respuesta correcta es 0 (por ejemplo, la predicción es 0,1), el costo es bajo porque en este caso la respuesta es 0. Si el modelo predice un resultado erróneo (por ejemplo, 0,9), el costo será elevado.

De hecho  $X=1$ , el costo es tan alto que aquí podemos recortado para apreciarlo más legible el gráfico

### ¿Por qué no MSE?

MSE y la pérdida de registros son métricas similares. La pérdida de registros castiga las respuestas incorrectas de manera más contundente que SE.



En este gráfico donde apreciamos la respuesta correcta es = 0, las predicciones por encima de 0.8, tiene un costo mucho mayor que la pérdida logarítmica que el MSE

#### Limitaciones de las funciones de costo

El uso de una función de costo única para la evaluación humana del modelo siempre está limitado porque no indica qué tipo de errores está cometiendo el modelo.

#### Mejora de los modelos de clasificación

Vamos a analizar con detalle las dos formas principales de mejorar el rendimiento del modelo de clasificación: proporcionar características adicionales y ser selectivos sobre lo que se incluye en el modelo

#### Proporcionar características adicionales

La regresión logística no tiene que limitarse a una sola entrada. Puede combinar características para realizar predicciones

#### Evitar el sobreentrenamiento

Si proporcionamos características adicionales que no son especialmente útiles, el modelo puede sobreentrenarse. Puede parecer que funciona mejor, pero realmente funciona peor

#### Evitar un entrenamiento deficiente

El uso de características de forma inocente también puede llevar a un entrenamiento deficiente y a no realizar predicciones de la manera correcta posible

## Módulo 6:

### Selección y personalización de arquitecturas e hiperparámetros mediante bosques aleatorios

- Detectar nuevos tipos de modelos: árboles de decisión y bosques aleatorios
- Saber cómo la arquitectura del modelo afecta al rendimiento
- Practicar el trabajo con hiperparámetros para mejorar la eficacia del entrenamiento

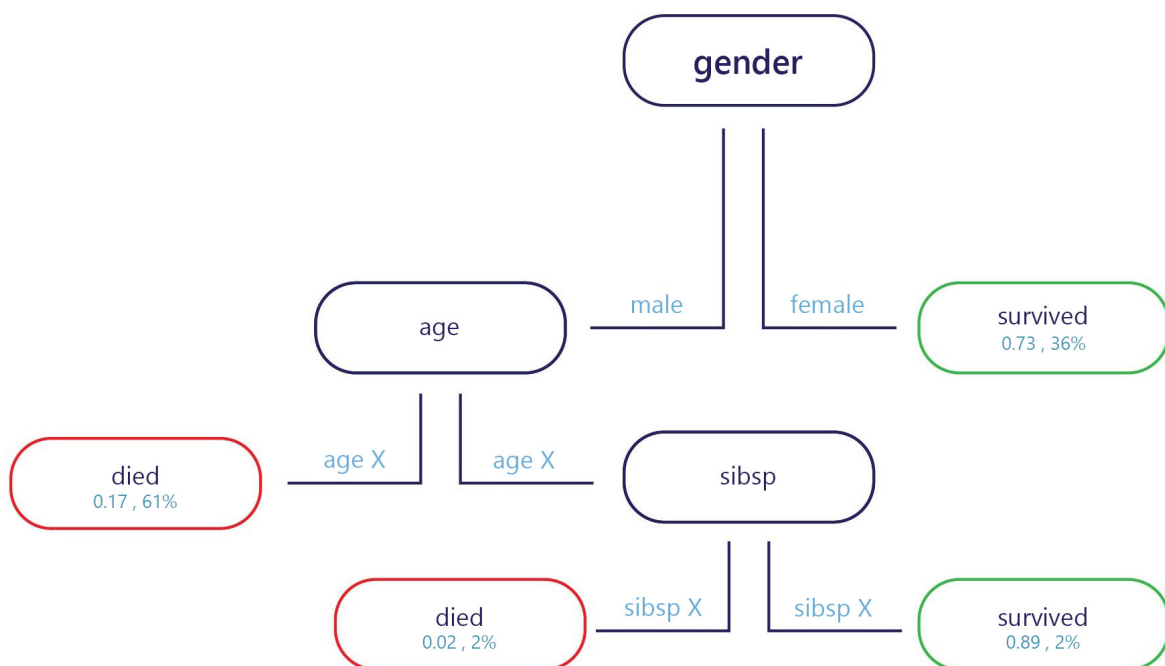
## ¿Árboles de decisión y arquitectura de modelos

En el aprendizaje automático, la arquitectura hace referencia a un concepto similar a lo que siempre creemos que es. ¿Cuántos parámetros tiene y cómo están vinculados para lograr un cálculo? ¿Realizamos muchos cálculos en paralelo(ancho) o tenemos operaciones en serie que se basan en un cálculo anterior (profundidad)?

Estas decisiones se suelen tomar antes de entrar al modelo, vamos a dar un ejemplo con los árboles de decisión

## ¿Qué es un árbol de decisión?

Es un árbol de flujos. Los árboles de decisión son un modelo de categorización que divide las decisiones en varios pasos



En cada nodo tenemos una condición "if" decide qué rama seguir, una vez que llega al final de las hojas se le asigna una etiqueta.

## ¿Cómo se entrenan los árboles de decisión?

En el primer nodo, se evalúa todo el conjunto de datos del entrenamiento. A partir de ahí se selecciona una característica que puede separar mejor en dos subconjuntos que tengan dos etiquetas homogéneas.

Ejemplo:

Peso (característica)	Edad (característica)	Ganó una medalla (etiqueta)
90	18	No
80	20	No
70	19	No
70	25	No
60	18	Sí
80	28	Sí
85	26	Sí
90	25	Sí



En este caso podemos separarlo en torno a la edad, teniendo en cuenta que la mayoría de los medallistas tienen más 24 años, esto nos dejaría en dos subconjuntos como los siguientes:

Subconjunto 1: (menores a 24 años)

Peso (característica)	Edad (característica)	Ganó una medalla (etiqueta)
90	18	No
80	20	No
70	19	No
60	18	Sí

Subconjunto 2: (mayores a 24 años)

Peso (característica)	Edad (característica)	Ganó una medalla (etiqueta)
70	25	No

80	28	Sí
85	26	Sí
90	25	Sí

Esto determina un porcentaje más detallado de los medallista que obtuvieron una medalla, pero para tener más detalle podríamos a partir de estos subconjuntos determinar por el peso y determinar otro resultado, esto es básicamente un árbol de decisión

#### ¿Puntos fuertes y débiles de los árboles de decisión

Se considera que los árboles tienen sesgos muy bajos. Esto significa que se buenos para identificar características que son importantes para etiquetar algo correctamente  
La principal debilidad es el sobreajuste

#### ¿La arquitectura del modelo afecta el sobreajuste

La forma en que estructuramos el árbol de decisión es la clave para evitar sus puntos débiles.

#### ¿Bosques aleatorios y selección de arquitecturas

La experimentación con arquitecturas suele centrarse en la creación de modelos eficaces. Y así lo hemos hecho a un nivel básico con árboles de decisión, pero el único límite es nuestra imaginación, y la memoria del equipo.

#### ¿Bosque aleatorio?

Básicamente es una colección de árboles de decisión que se usan juntos para calcular qué etiqueta se debe asignar a nuestra muestra.

#### ¿Cómo se entrena un bosque aleatorio?

Se construye con la idea que una árbol de decisión es muy sesgado o sobre ajustado, para ello se tiene que entrenar varios árboles de decisión con ligeros ajustes y así diferenciar los sesgos que tiene cada uno.

#### ¿Ventajas de un bosque aleatorio

Se suele comparar con las redes neuronales, que son otro tipo de modelo conocido y de alto rendimiento. A diferencia de las redes neuronales, los bosques aleatorios son fáciles de entrenar: no necesitan un conjunto de datos para un buen rendimiento en el entrenamiento

#### Desventajas de un bosque aleatorio

Son difíciles de comprender, aunque estos modelos son extremadamente transparentes(c/árbol se puede inspeccionar y entender), a menudo tiene tantos árboles que resulta prácticamente imposible hacerlos.

#### Personalizar esta arquitectura

Lo más fácil es tener en cuenta el tamaño del bosque: el número de árboles que intervienen, junto el tamaño de los árboles

#### Hiperparametros en la clasificación

Se pueden considerar los valores que se usan para el entrenamiento. Por ejemplo, podríamos decidir si entrenamos de manera rápida o lenta. Normalmente, experimentamos con hiperparametros para optimizar el rendimiento de nuestro modelo.

#### Bosques aleatorios como ejemplo

Estos tienen diferentes tipos de hiperparametros disponibles.

#### Criterios de división

Durante el entrenamiento, el optimizador debe decidir cuándo dividir un nodo. Los métodos comunes de dividir los nodos se basan en la teoría de la información. A grandes rasgos, se podría decir considerar como dividir una muestra para que los dos submuestras resultantes sean “más puras” que la original

#### Disminución mínima de impurezas

Significa que un nodo se puede dividir si mejora el modelo algo o bastante

#### Número máximo de características

Lo importante es que cada árbol pueda recibir diferentes colecciones de características. Por ejemplo: Árbol 1: usa los parámetros “peso” y “altura”; Árbol 2: “altura” y “edad”

## Módulo 7:

### Matriz de confusión y desequilibrios de datos

- Evaluación del rendimiento de los modelos de clasificación
- Revisión de las métricas para mejorar los modelos de clasificación
- Mitigación de problemas de rendimiento de desequilibrio de datos

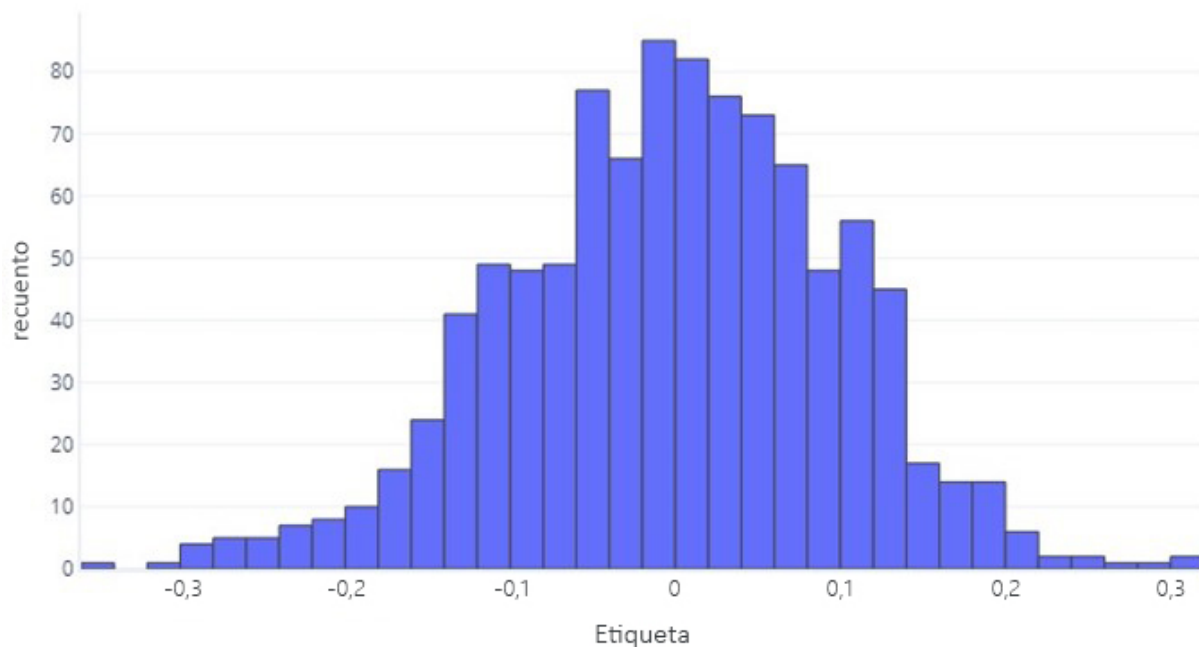
#### Matrices de confusión

Los datos se pueden considerar continuos, categóricos u ordinales(categóricos pero con un orden).Las matrices de confusión son un medio de evaluar el rendimiento de un modelo categórico.

Repaso antes de ver matrices de confusión

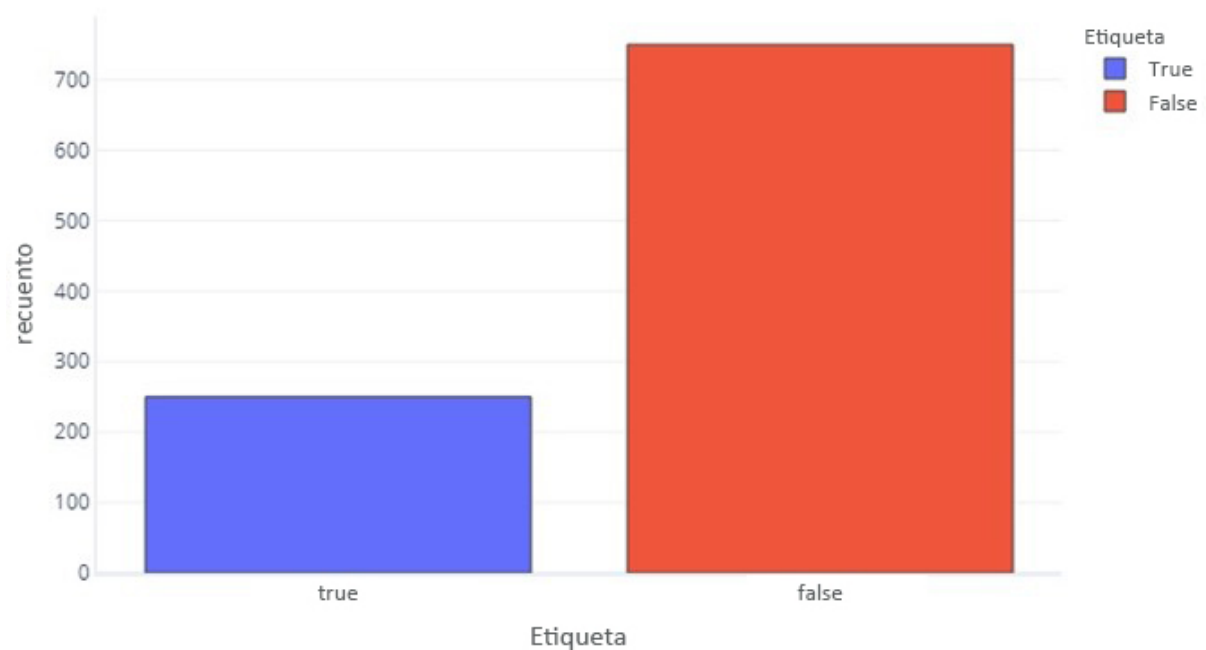
#### Distribuciones de datos continuos

Cuando queremos comprender los datos continuos, el primer paso suele ser ver cómo se distribuyen los datos. Miremos el siguiente histograma  
Podemos ver que la etiqueta es, por término medio, aproximadamente cero y que la mayoría de los puntos de datos se encuentran en -1 y 1. Aparece como simétrico



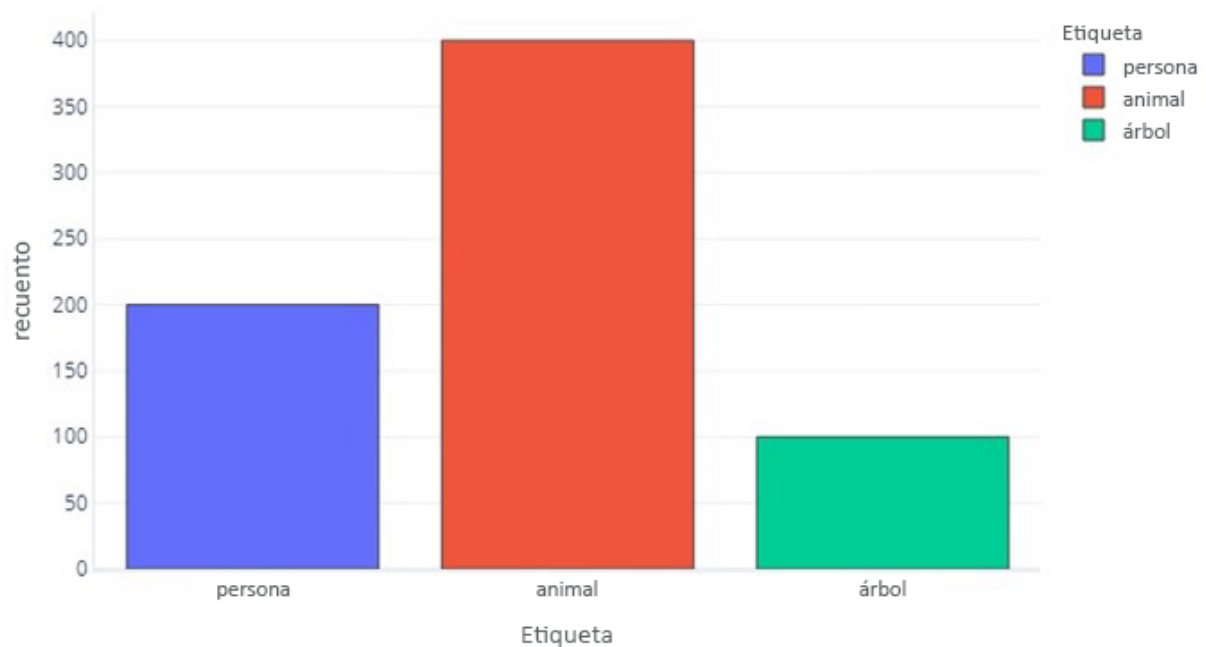
#### Distribuciones de datos categóricos

En algunos aspectos, los datos no son tan diferentes de los datos continuos. Una etiqueta binaria (verdadera/falso) podría aparecer de esta frecuencia como esta:



Esto nos indica que hay 750 muestras con la etiqueta “false” y 250 con la etiqueta “true”

Una etiqueta para tres categorías es parecida:



Esto nos indica que hay 200 etiquetas de “personas”, 400 de “animal” y 100 de “árbol”

La clave para comprender el rendimiento del modelo es combinar la tabla de predicción del modelo con la tabla de etiquetas de datos reales:

		Etiqueta real	
		Falso	Verdadero
		750	250
Predicción de modelos	Falso	700	TN
	Verdadero	300	FP
			FN
			TP

Cada celda de la matriz de confusión nos indica algo sobre el rendimiento del modelo.

### Verdaderos Negativos(TN)

El valor superior izquierdo mostrará el número de veces que el modelo predijo “false” y la etiqueta real fue también “false”; En otras palabras, se muestra cuántas veces el modelo predijo correctamente “false”. Supongamos que, en nuestro ejemplo, esto ha ocurrido 500 veces:

		Etiqueta real	
		Falso	Verdadero
		750	250
Predicción de modelos	Falso	700	500
	Verdadero	300	

### Falso Negativos(FN)

El valor superior derecho indica cuántas veces el modelo predijo “false” y la etiqueta real fue “true”. Ahora sabemos que son 200 ¿Cómo lo hago? Dado que el modelo predijo “false” 700 veces y 500 de esas veces lo hizo correctamente. Por lo tanto, 200 veces debe haber predicho “false” cuando no debería haberlo hecho.

		Etiqueta real	
		Falso	Verdadero
		750	250
Predicción de modelos	Falso	700	500
	Verdadero	300	200

### ─Falsos positivos(FP)

El valor inferior izquierdo, nos indica cuántas veces predijo el modelo “true”, pero la etiqueta real era “false”. Ahora sabemos que son 250, porque hubo 750 veces que la respuesta correcta fue “false”. 500 de estas veces aparecen en la celda superior izquierda(TN)

		Etiqueta real	
		Falso	Verdadero
		750	250
Predicción de modelos	Falso	700	500
	Verdadero	300	250

### ─Verdaderos positivos(TP)

Este es el número de veces que el modelo predijo correctamente “true”. Sabemos que son 50 por dos motivos. En primer lugar, el modelo predijo “true” 300, pero 200 veces el modelo predijo “false”.

		Etiqueta real	
		Falso	Verdadero
		750	250
Predicción de modelos	Falso	700	500
	Verdadero	300	250

Finalmente quedaría de esta manera:

		Etiqueta real	
		Falso	Verdadero
Predicción de modelos	Falso	500	200
	Verdadero	250	50

Hemos coloreado las celdas para resaltar cuándo el modelo realizó predicciones correctas. A partir de esto, sabemos no solo la frecuencia con la que el modelo realizó determinado tipos de predicciones, sino también la frecuencia con la que fueron correctas o incorrectas

Con más etiquetas quedaría así:

		Etiqueta real		
		Persona	Animal	Árbol
Predicción de modelos	Persona	500	200	50
	Animal	250	50	20
	Árbol	10	150	100

Cuando hay tres categorías, ya no se aplican métricas como TP, pero todavía podemos ver exactamente con qué frecuencia el modelo ha cometido errores

#### Desequilibrios de los datos

Cuando nuestras etiquetas tienen más de una categoría que otras, se dice que hay desequilibrio de datos.



Por ejemplo, en el script anterior , intentamos identificar los objetos que han encontrado los sensores de drones. Nuestros datos están desequilibrados porque hay un número enorme de diferentes excursionistas, animales, árboles y rocas en nuestros datos de entrenamiento.

¿Por qué son importantes los desequilibrios de datos?

Porque los modelos pueden imitar estos desequilibrios cuando es deseable. Por ejemplo, imagine que entrenamos un modelo de regresión logística para identificar objetos como excursionista o no excursionistas. Si en los datos de entrenamiento dominan la etiqueta “excursionista” el entrenamiento se sesgará el modelo devolverá constantemente esta etiqueta.

Evitar las consecuencias de los datos desequilibrados

- Realizar una mejor selección de datos
- “Nuevo muestreo” de los datos para que contengan duplicados de las etiquetas minoritarias
- Realizar cambios en la función de costo para que dé prioridad a las etiquetas menos comunes.

## Módulo 8:

### Medición y optimización del rendimiento del modelo con ROC y AUC

- Comprender cómo crear curvas de ROC
- Explorará cómo evaluar y comparar modelos mediante estas curvas
- Practicará un ajuste a un modelo mediante características trazadas en curvas de ROC

Análisis de la clasificación con curvas de características operativas del receptor (ROC)

El ajuste del modelo en este punto puede ser complejo, pero podemos usar un sencillo paso final para mejorar aún más el buen funcionamiento de nuestro modelo, Sin embargo, para entender, es necesario volver a aspectos básicos.

Probabilidades y categorías

Muchos modelos tienen fases de toma de decisiones y la última suele ser simplemente un paso de binarización.

Durante la binarización, las probabilidades se convierten en una etiqueta dura.

Supongamos que el modelo se proporciona con características y calcula probabilidades del 75% que se mostrará un excursionista y del 25% de un árbol; Un objeto no puede ser 75% algo y el 25% otro , en conclusión en la binarización es uno o otro. El mismo modelo aplica un umbral del 50% , este es un idea que el modelo va a seguir pero si el modelo está sesgado va a tender a elegir uno más que el otro y el umbral llegaría a cambiar

Actualizador en función de las matrices de decisión

Características populares:

- Tasa de TP(sensibilidad): la frecuencia con las etiquetas “True” se identifica correctamente como “True”. Por ejemplo, la frecuencia con la que el modelo predice “excursionistas” cuando lo que se muestra es un excursionista
- Tasa de FP(índice de falsas alarmas): la frecuencia con las etiquetas “False” se identifican incorrectamente como “True”. Por ejemplo, la frecuencia con la que el modelo predice “excursionistas” cuando se muestra árbol

Estos exámenes no permiten comprender el rendimiento de un modelo

#### ¿Curvas ROC

Son un gráfico en el que se trata de verdaderos positivos frente a las tasas de falsos positivos

#### ¿ROC buenas, ROC malas

Comprender las curvas de ROC buenas y malas es algo que se hace mejor en un entorno interactivo

#### ¿Comparación y optimización de curvas ROC

Las curvas de características operativas del receptor (ROC) nos permiten comparar modelos entre sí y ajustar nuestro modelo seleccionado

#### ¿Ajustes de un modelo

El uso más obvio de una curva de ROC es elegir un umbral de decisión que proporcione el mejor rendimiento

Por lo general, no hay un solo umbral que proporcione la mejor tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) más baja. Esto significa que el umbral óptimo depende de lo que se intente lograr

#### ¿Comparación de modelos con AUC

Puede usar las curvas ROC para comparar modelos entre sí, al igual que la funciones costo. Al fin y al cabo, lo más importante de un modelo es qué rendimiento tendrá en el mundo real.

#### ¿Cómo se comparan los modelos ROC?

La manera más fácil de comparar las ROC numéricamente es usar el área bajo la curva (AUC). Literalmente, se trata del área del gráfico que se encuentran por debajo de la curva