

Árboles de Decisión

Delta Analytics construye capacidad técnica alrededor del mundo.



El contenido de este curso está siendo desarrollado activamente por Delta Analytics, una organización sin fines de lucro 501(c)3 del Área de la Bahía que apunta a capacitar a las comunidades para aprovechar sus datos.

Por favor comuníquese con cualquier pregunta o comentario a inquiry@deltanalytics.org.

Descubre más sobre nuestra misión [aquí](#).

Módulo 4.1:

Árboles de decisión



Checklist del módulo

- ❑ Árboles de decisión
 - ❑ Intuición
 - ❑ La "mejor" división
 - ❑ Modelo de rendimiento
 - ❑ Optimización del modelo (poda)



¿Dónde estamos?

Hasta este punto ya hemos revisado el algoritmo de regresión lineal, regresión logística, naive bayes y SVM.

*En este módulo vamos a introducir un nuevo algoritmo. **Árboles de decisión** son otro tipo de algoritmo que puede lograr el mismo objetivo de predicción que los otros algoritmos.*

También profundizaremos en los pros y los contras de los árboles de decisión en este módulo.



¿Dónde estamos?

Algoritmos
Supervisados

Regresión
Lineal

Regresión
Logística

SVM

Árboles de
Decisión

Algoritmos
de
Conjunto

Hagamos una introducción
rápida a los árboles de
decisiones y conjuntos



Hoy, comenzaremos revisando el algoritmo de árbol de decisiones.

Un árbol de decisión es un conjunto de reglas que podemos usar para clasificar datos en categorías (**también se puede usar para tareas de regresión**).

Los humanos a menudo usan un enfoque similar para llegar a una conclusión. **Por ejemplo, los médicos hacen una serie de preguntas para diagnosticar una enfermedad. El objetivo de un médico es hacer la cantidad mínima de preguntas necesarias para llegar al diagnóstico correcto.**



¿Qué le pasa a mi paciente?

Ayúdame a armar algunas preguntas que pueda usar para diagnosticar a los pacientes correctamente.

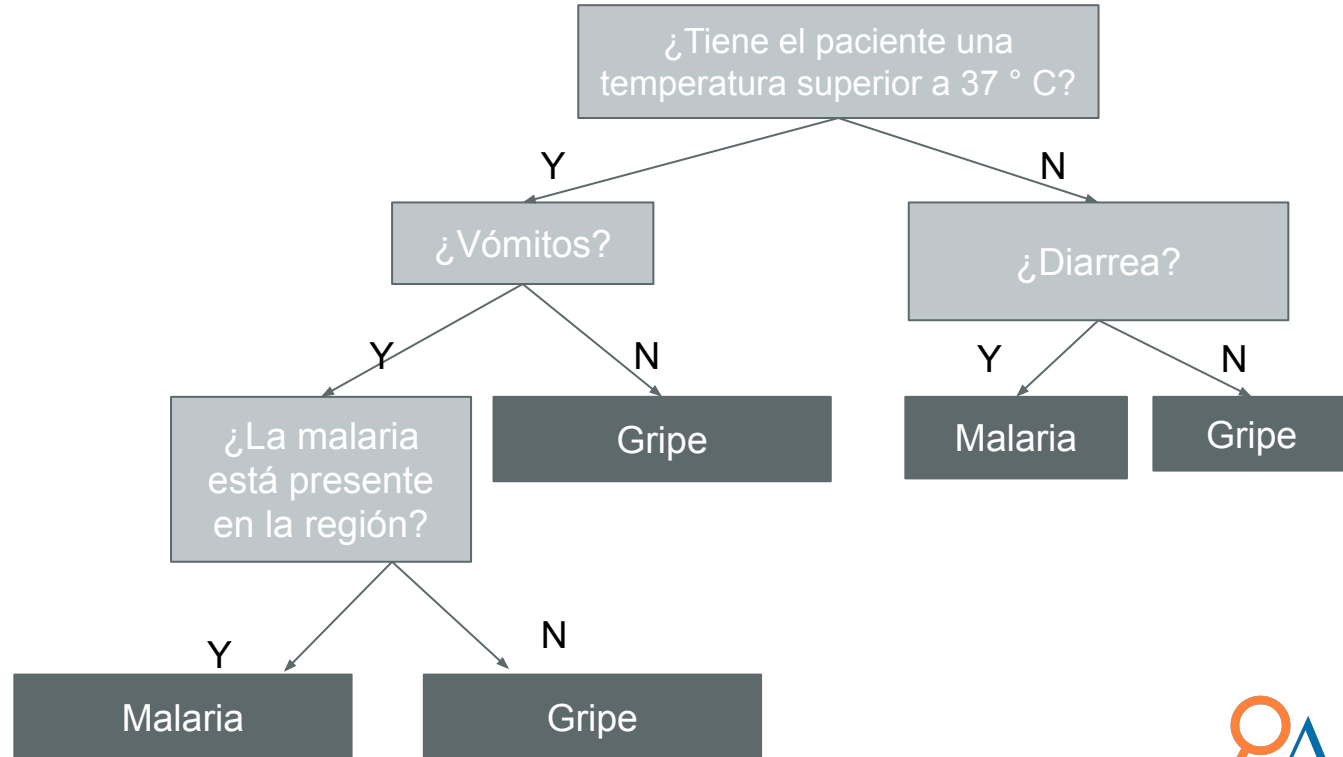
Árbol de decisión

Los árboles de decisión son intuitivos porque son similares a cómo tomamos muchas decisiones.



Mi árbol de decisión de diagnóstico mental podría verse más o menos así.

¿Cómo es la forma en que pienso sobre esto diferente de un algoritmo de aprendizaje automático?



Los árboles de decisión son intuitivos y pueden manejar relaciones más complejas que la regresión lineal.

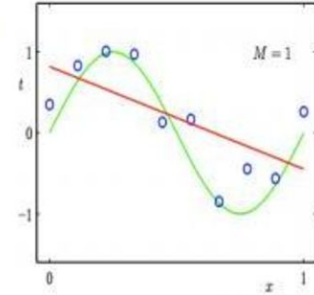
Una regresión lineal es una línea de tendencia global única.

Esto lo hace inflexible para relaciones más sofisticadas.

Sources:

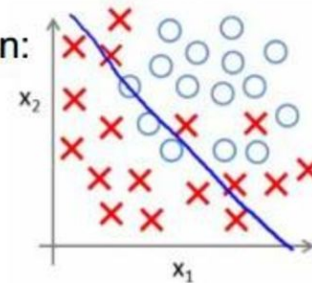
<https://www.slideshare.net/ANITALOKITA/winnow-vs-perceptron>,
<http://www.turingfinance.com/regression-analysis-using-python-statmodels-and-quandl/>

Regression:

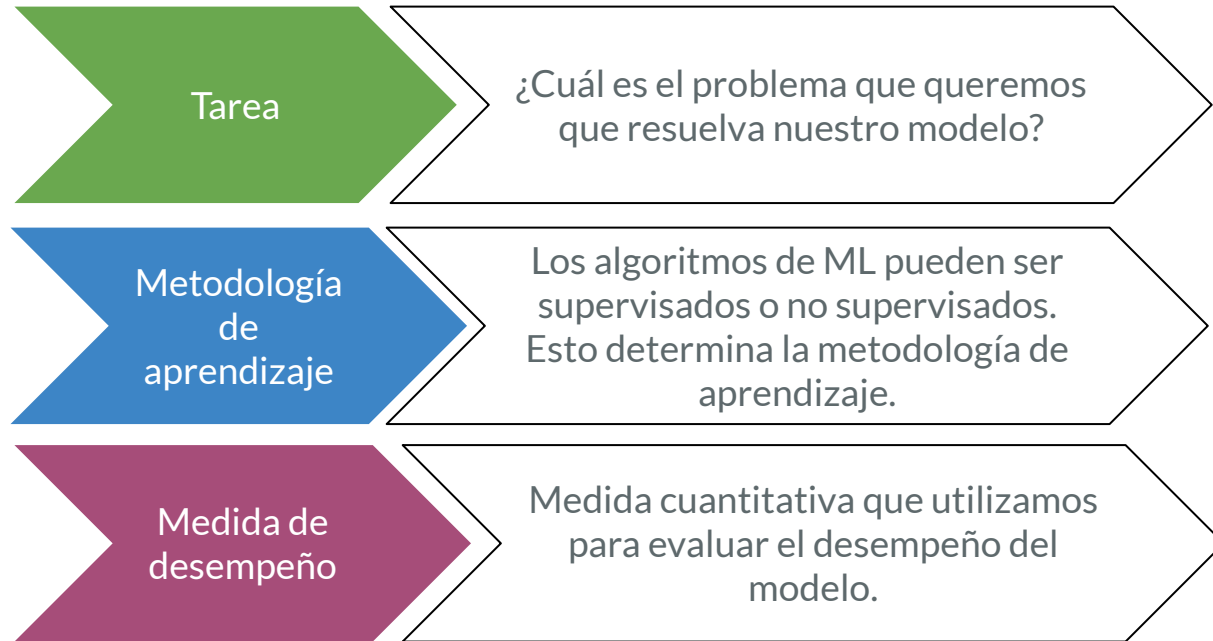


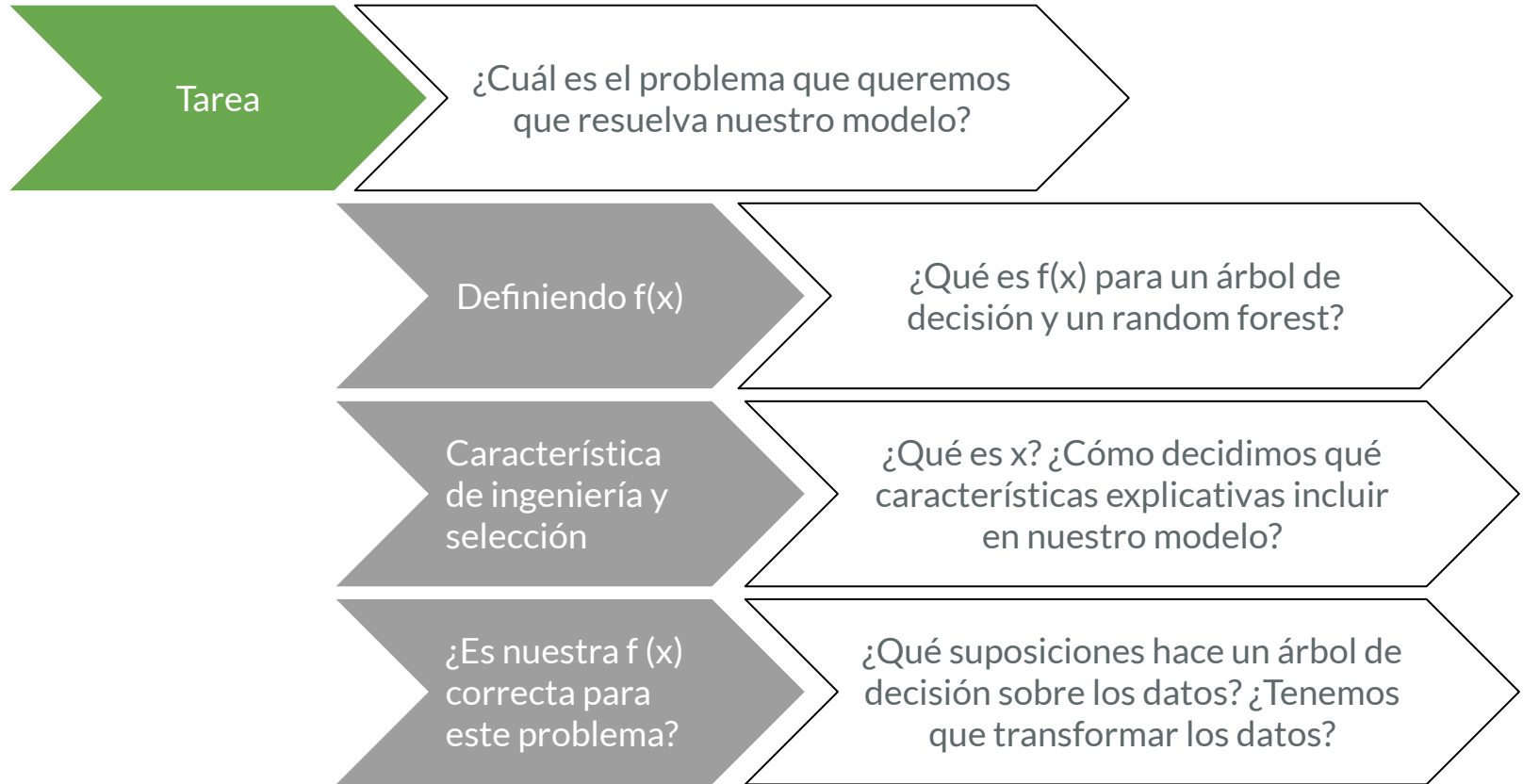
predictor too inflexible:
cannot capture pattern

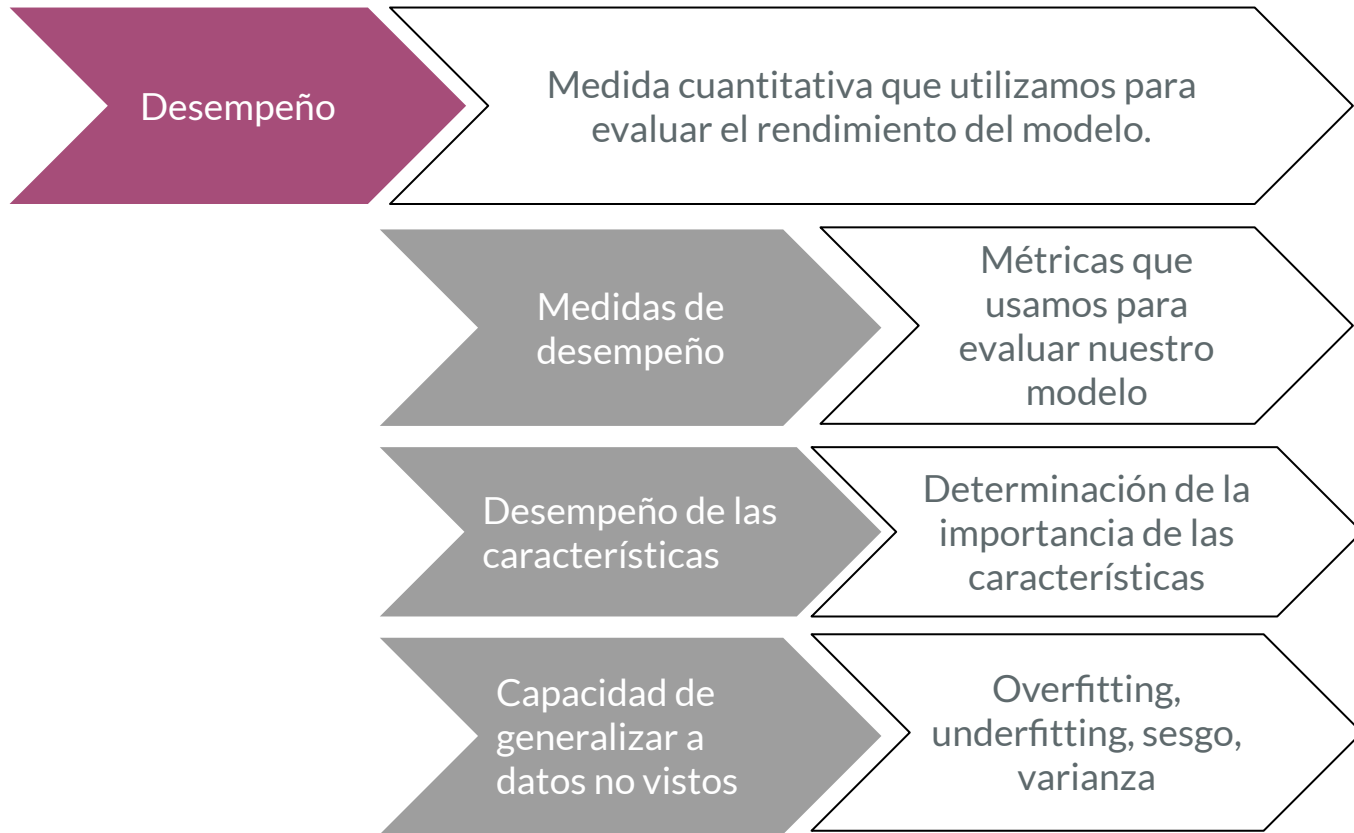
Classification:



Usaremos nuestro marco ahora familiar para discutir tanto los árboles de decisión como los algoritmos de conjunto:







Metodología
de
aprendizaje

Los modelos de árbol de decisión son supervisados. ¿Cómo afecta eso al proceso de aprendizaje?

¿Cómo aprende nuestro modelo ML?

Descripción general de cómo se enseña el modelo.

¿Cuál es nuestra función de pérdida?

Cada modelo supervisado tiene una función de pérdida que quiere minimizar.

Proceso de optimización

¿Cómo minimiza el modelo la función de pérdida?

Árbol de decisión



Árbol de decisión:

Pros

- Imita la intuición humana (¡tomamos decisiones de la misma manera!)
- No *prescriptivo* (es decir, los árboles de decisión no asumen una distribución normal)
- Se puede entender e interpretar intuitivamente como un diagrama de flujo o una división del espacio de características.
- Puede manejar datos no lineales (no es necesario transformar los datos)

Cons

- **Susceptible a sobreajuste** (bajo rendimiento en el conjunto de prueba)
- Alta varianza entre conjuntos de datos.
- Puede volverse inestable: pequeñas variaciones en los datos de entrenamiento generan árboles completamente diferentes.



Tarea



Tarea

¿Qué estamos prediciendo?



¿Cómo diagnostica un médico a sus pacientes en función de sus síntomas?

Tarea

Un médico elabora un diagnóstico a partir de sus síntomas.



Comienzo preguntando a los pacientes una serie de preguntas sobre sus síntomas.

Utilizo esa información para construir un diagnóstico.

temperatura	vomitos	Escalofríos	Diagnóstico	Diagnóstico predicho
X1	X2	X3	Y	Y*
39.5°C	Si	Severos	Gripe	Gripe
37.8°C	No	Severos	Malaria	Malaria
37.2°C	No	Leves	Gripe	Malaria
37.2°C	Si	No	Gripe	Gripe

Hay algunas preguntas que el médico no necesita hacer porque lo aprende de su entorno. (Por ejemplo, un médico sabría si la malaria está presente en esta región). Sin embargo, una máquina tendría que aprender explícitamente esta característica.

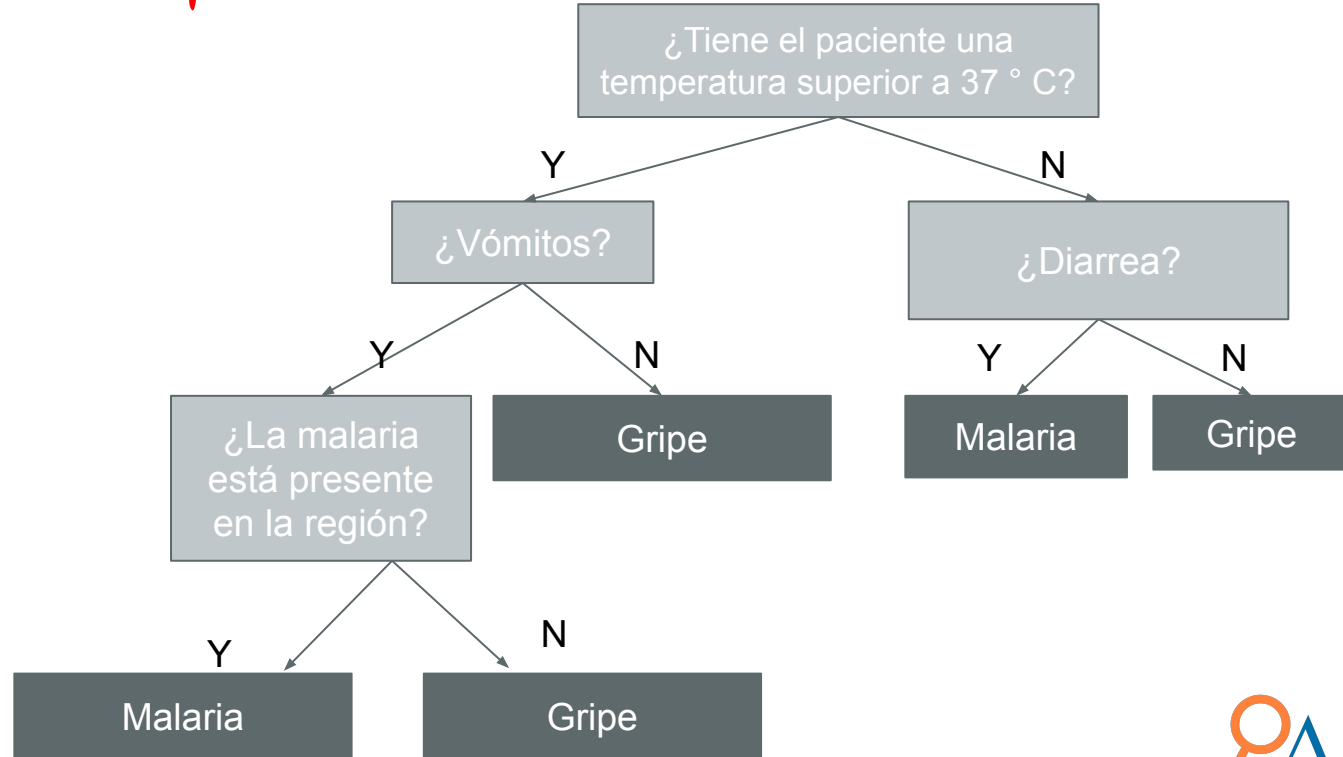


Tarea

Un médico hace un camino de diagnóstico mental para llegar a su conclusión. Ella está construyendo un árbol de decisión!



¿Cómo se compara mi manera de decidir el valor de la división con un algoritmo de aprendizaje automático?



Intuición Humana



Según mi experiencia como médico, sé que hay ciertas preguntas cuyas respuestas separan rápidamente la gripe de la malaria.



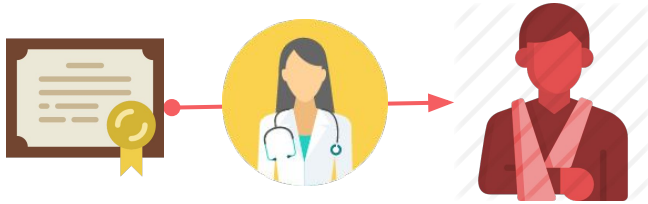
Árbol de Decisión

En cada división, podemos determinar la mejor pregunta para maximizar el número de observaciones correctamente agrupadas en la categoría correcta.

- Tanto el médico como el árbol de decisiones intentan llegar al número mínimo de preguntas (divisiones) que deben pedirse para diagnosticar correctamente al paciente.
- Diferencia clave: cómo se determina el orden de las preguntas y el valor de división. Un médico hará esto en función de la **experiencia**, un árbol de decisión **formalizará** este concepto utilizando una función de pérdida.

Tarea

Machine learning agrega el poder de los datos a la tarea de nuestro médico.

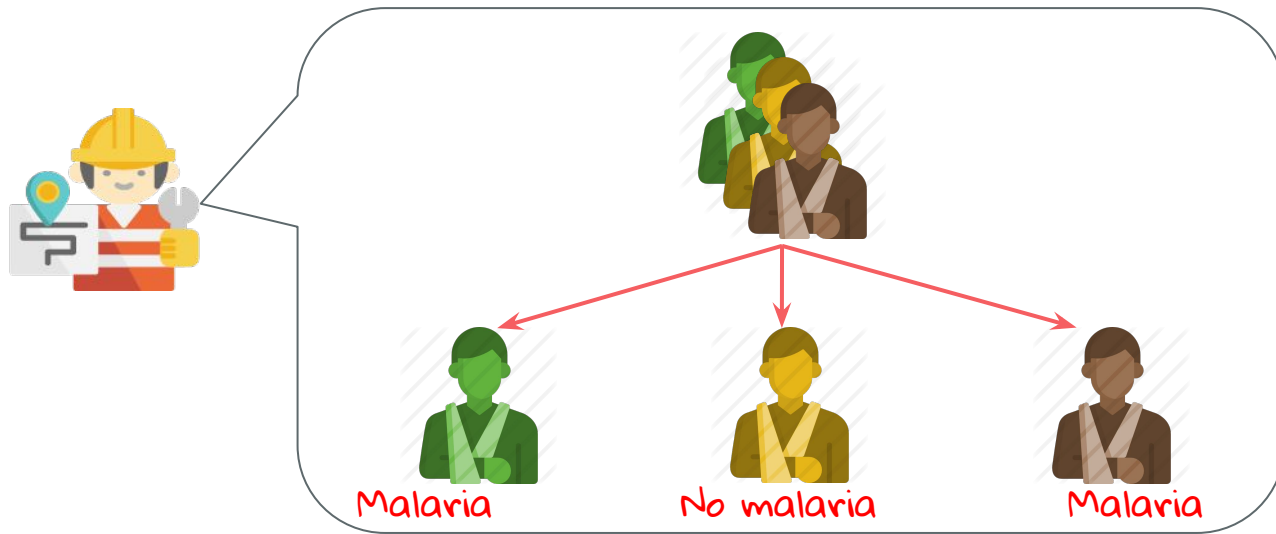


Para determinar si un nuevo paciente tiene malaria, un médico usa su experiencia y educación, y el machine learning crea un modelo con datos.



Al usar un modelo de aprendizaje automático, nuestro médico puede usar tanto su experiencia como sus datos para tomar la mejor decisión.

Los árboles de decisión actúan de la misma manera que la categorización humana, ellos (dividen) los datos basados en las respuestas a las preguntas.



¿A dónde voy?

Mr. Model crea un diagrama de flujo (el modelo) separando todo nuestro conjunto de datos en distintas categorías. Una vez que tengamos este modelo, sabremos en qué categoría se encuentra nuestro nuevo paciente.



Para tener una idea de cómo funciona esta "división", juguemos un juego de adivinanzas. **Estoy pensando en algo que es azul.**

Puede hacer 10 preguntas para adivinar de qué se trata.



Las primeras preguntas que probablemente haría podrían incluir:

- ¿Está vivo? (No)
- ¿Está en la naturaleza? (Sí)



Luego, a medida que se acerque a la respuesta, sus preguntas se volverán más específicas.

- ¿Es sólido o líquido? (Líquido)
- ¿Es el **océano**? (¡Sí!)

Este proceso es tu árbol de decisión mental. De hecho, esta estrategia captura muchas de las cualidades de nuestro algoritmo!



Algunas estrategias ganadoras incluyen:

- Use preguntas desde el principio que eliminen la mayor cantidad de posibilidades.
- Las preguntas a menudo comienzan de manera amplia y se vuelven más granulares.

Creaste tu propio árbol de decisión **basado en tu propia experiencia de lo que sabes que es azul** para hacer una conjetura sobre lo que yo estaba pensando (**el océano**).

Del mismo modo, un modelo de árbol de decisión hará una conjetura, pero en lugar de usar la experiencia, usa datos.

Ahora, construyamos un vocabulario formal para discutir los árboles de decisión.

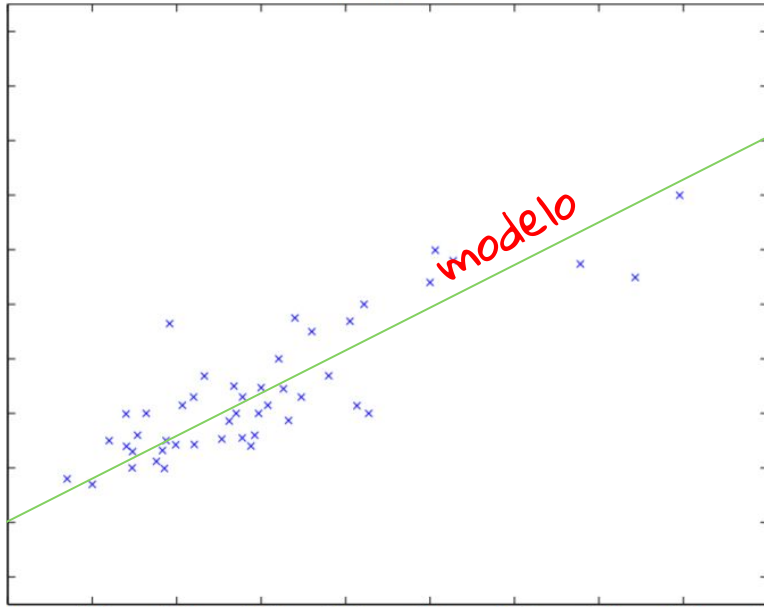


Árbol de
decisiones

Definiendo
 $f(x)$

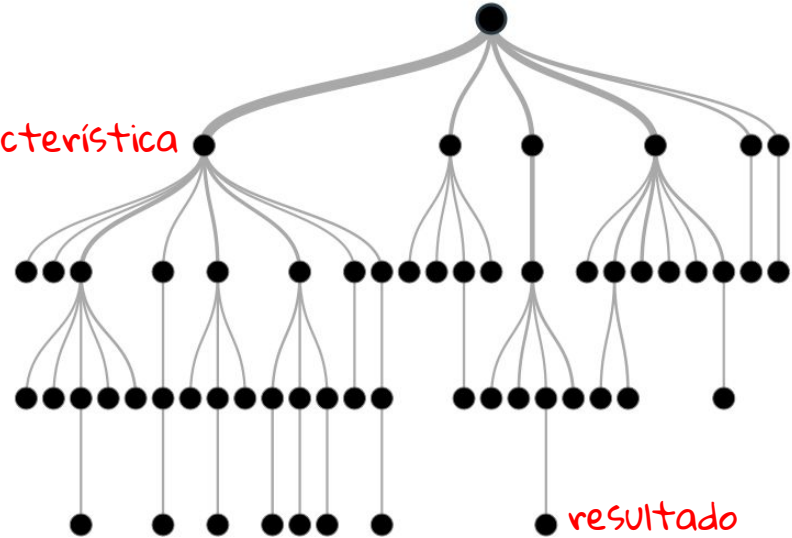
Al igual que una regresión lineal, un árbol de decisión tiene características explicativas y un resultado. Nuestra $f(x)$ es la ruta de decisión desde la parte superior del árbol hasta el resultado final.

resultado



Característica

característica



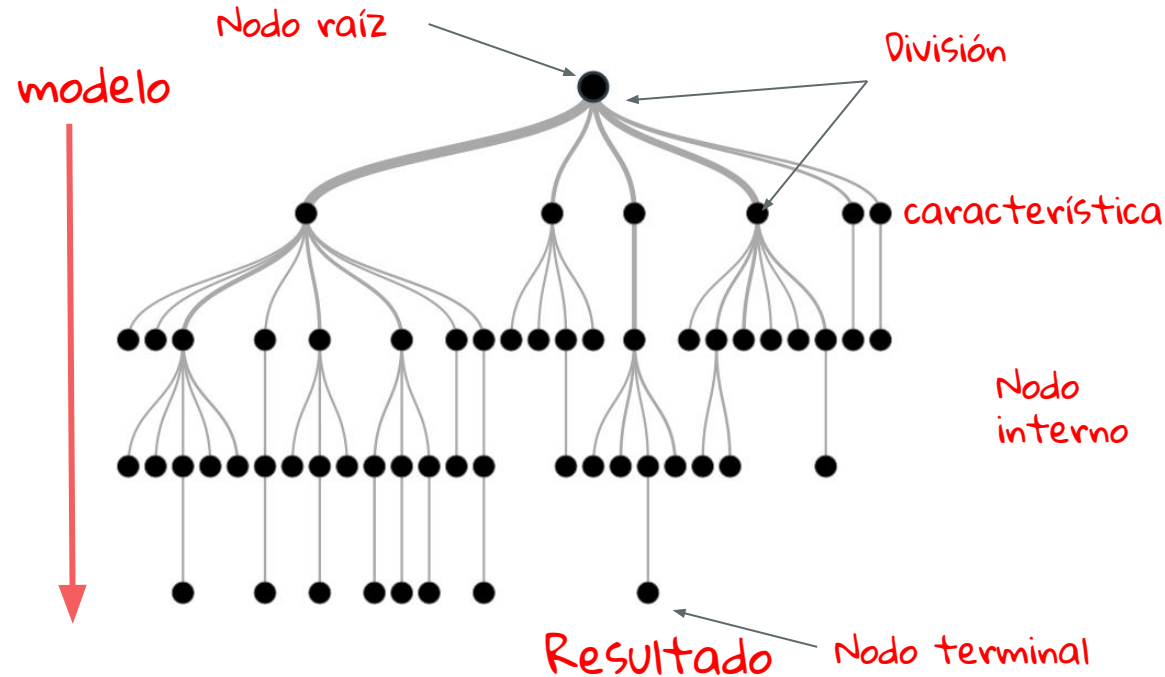
modelo

resultado



Árbol de
decisiones

Nuevo
vocabulario



División

Las "decisiones" del modelo del árbol de decisiones. El modelo decide qué características usar para dividir los datos.

Nodo raíz

Nuestro punto de partida. La primera división se produce aquí a lo largo de la característica que el modelo decide que es **más importante**.

Nodo interno

Divisiones intermedias. Corresponde a un subconjunto de todo el conjunto de datos.

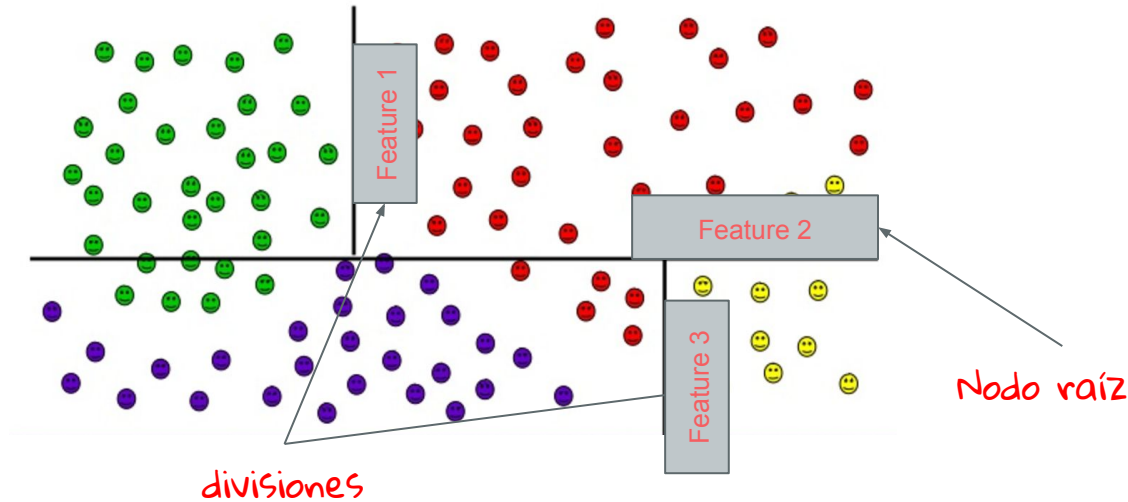
Nodo terminal

Resultado pronosticado. Corresponde a un subconjunto más pequeño de todo el conjunto de datos.

Árbol de
decisiones

Nuevo
vocabulario

Un árbol de decisión divide el espacio de características (el conjunto de datos) a lo largo de las características. El orden de las divisiones está determinado por el algoritmo.



Esta visualización de divisiones en el espacio de características es una forma diferente pero igualmente precisa de conceptualizar los árboles de decisión como el diagrama de flujo en la diapositiva anterior.

Tarea

Alguna
terminología
adicional

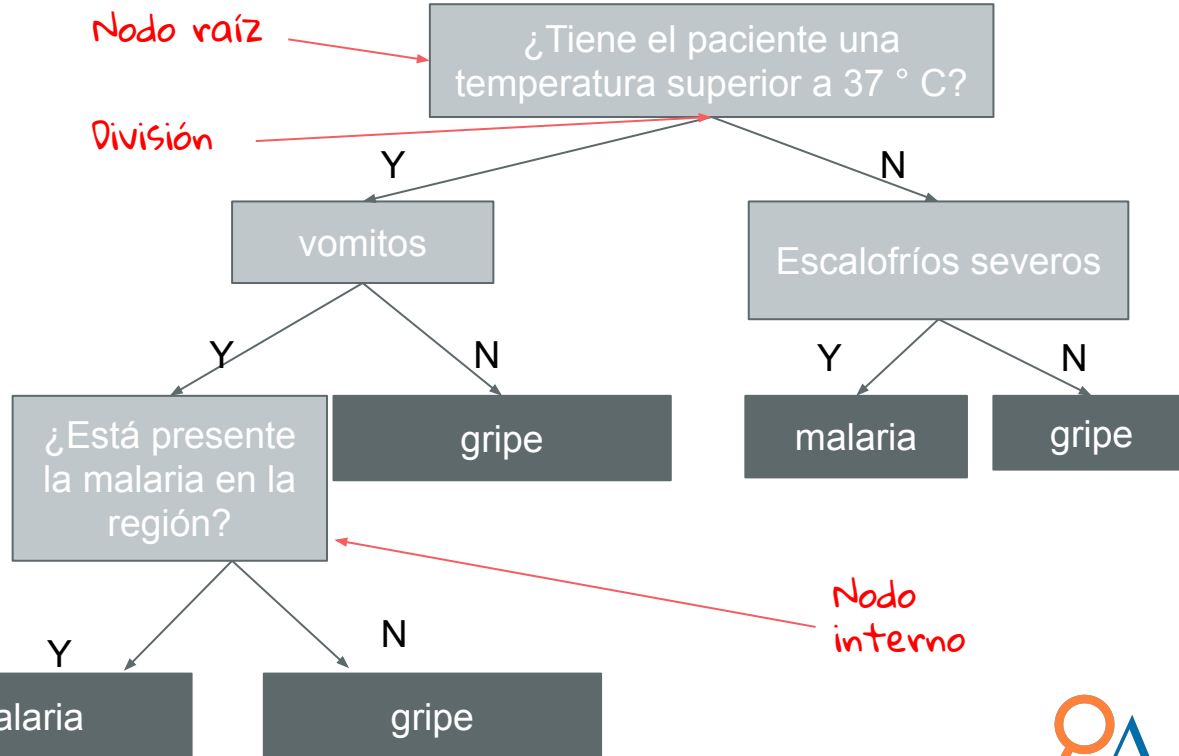
Las divisiones en la parte superior del árbol tienen una importancia descomunal en el resultado final.

Nodo raíz

División



¿Cuál es el nodo raíz, división, nodo interno y nodo terminal en este ejemplo?



Nodo
interno

Nodo terminal



Árbol de
decisión

Definiendo
 $f(x)$

Veamos otra pregunta para combinar nuestra intuición con el vocabulario formal. ¿Qué debería hacer Sam este fin de semana?



¡Comencemos
este fin de
semana!

Y = Elección de actividad de fin de semana

Bailar

Cocinar la cena en casa

Comer en un restaurante elegante

Concierto de música

Caminar en el parque

Conjunto de datos de entrenamiento:

Tienes un conjunto de datos de lo que ella ha hecho todos los fines de semana durante los últimos dos años.



¿Qué hará Sam este fin de semana?

Tiene
pareja

X1

Yes

Padres en
la ciudad

X2

No

Ahorros
(\$)

X3

\$80

¿Qué hará Sam?

Y

Bailar

Cocinar la cena en casa

Comer fuera

Concierto de música

Caminar en el parque

Caminar en el parque

Bailar

Comer fuera



Tenemos datos históricos sobre lo que Sam ha hecho los fines de semana pasados (Y), así como algunas variables explicativas.

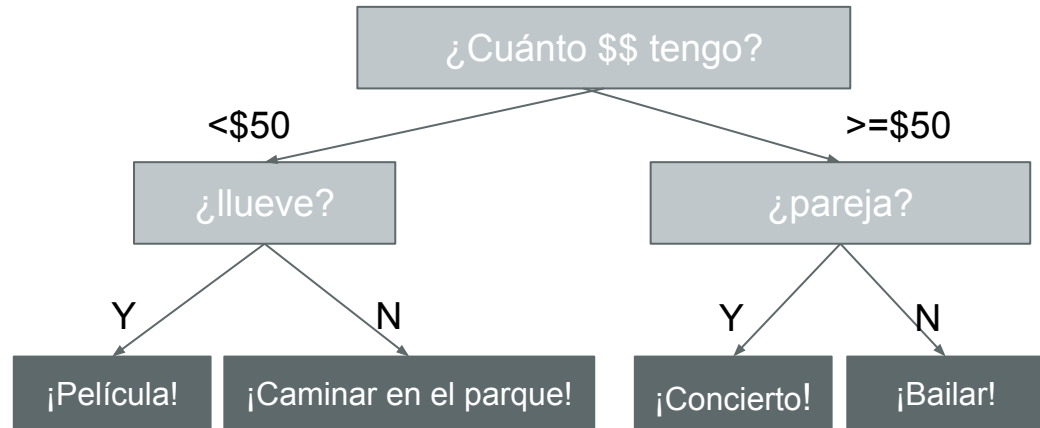
Tarea del
árbol de
decisiones

Definiendo
 $f(x)$

El árbol de decisión $f(x)$ predice el valor de una variable objetivo al aprender reglas de decisión simples inferidas de las características de los datos.

En este ejemplo, predecimos la actividad de Sam durante el fin de semana utilizando reglas de decisión capacitadas sobre el comportamiento histórico del fin de semana.

Nuestra característica predictiva más importante es el presupuesto de Sam. ¿Cómo sabemos esto? Porque es el nodo raíz.



Tarea del
árbol de
decisiones

$f(x)$ como una
función
espacial

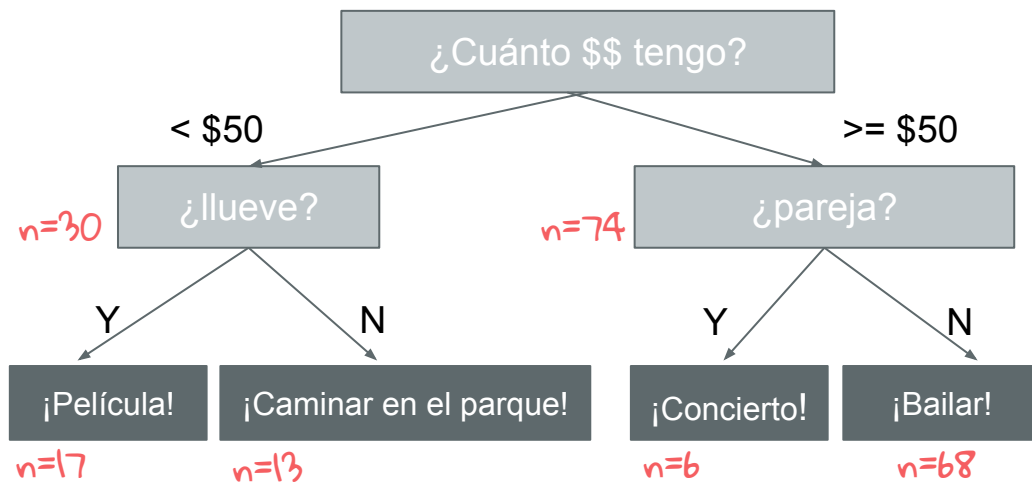
Un ejemplo de cómo funciona el
algoritmo:



$n=104$

Un árbol de decisión divide el conjunto de datos ("**espacio de características**"). Aquí, el conjunto de datos completo = datos de 104 fines de semana. Puedes ver como cada división forma conjuntos de datos en grupos más pequeños progresivamente.

Ahora, cuando queremos predecir lo que sucederá este **fin de semana**, ¡podemos hacerlo!

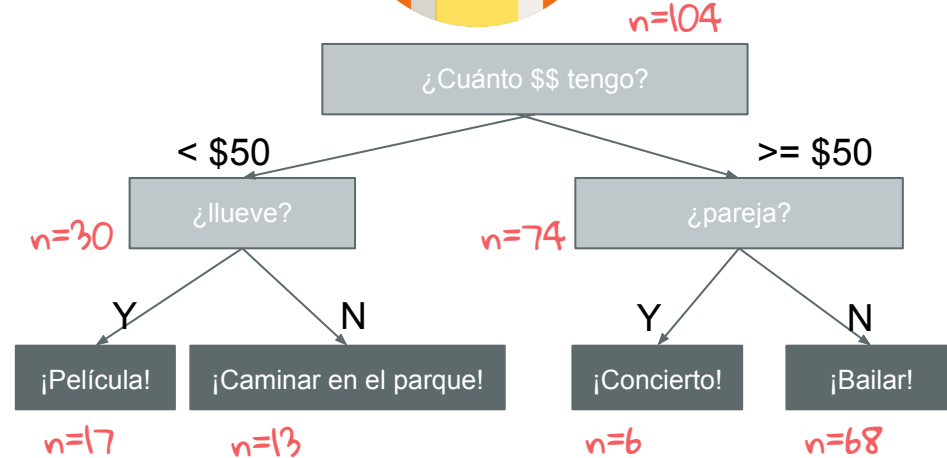
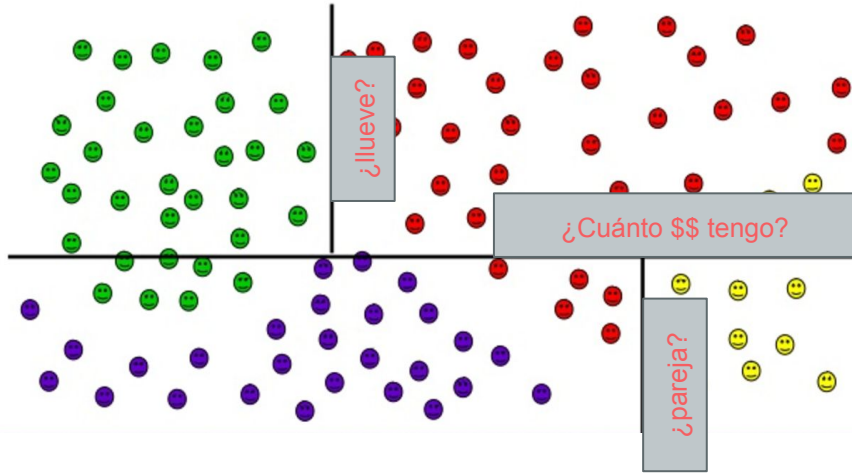


Tarea del
árbol de
decisiones

$f(x)$ como una
función
espacial

Cada una de las cuatro opciones de fin
de semana están agrupadas
espacialmente por nuestras divisiones.

Visualized in the data:



Ahora entendemos la mecánica de la tarea del árbol de decisiones. Pero, **¿cómo aprende un árbol de decisión?**
¿Cómo sabe dónde dividir los datos y en qué orden?

Pasamos ahora a examinar la metodología de aprendizaje de los árboles de decisión.



Metodología de aprendizaje



Recuerda que la diferencia clave entre un médico y nuestro modelo es cómo se determina el orden de las preguntas y el valor dividido.

Intuición Humana



Según mi experiencia como médico, sé que hay ciertas preguntas cuyas respuestas separan rápidamente la gripe de la malaria.



Árbol de Decisión

En cada división, podemos determinar la mejor pregunta para maximizar el número de observaciones agrupadas correctamente en la categoría correcta.

La diferencia clave entre un médico y nuestro modelo es cómo se determina el orden de las preguntas y el valor dividido.



Las dos palancas clave que se pueden cambiar para mejorar la precisión de nuestro diagnóstico médico son:

- El orden de las preguntas.
- El valor dividido en cada nodo. (Por ejemplo, el límite de temperatura)

Un médico tomará estas decisiones basándose en la experiencia, un árbol de decisiones establecerá estos **parámetros** para minimizar nuestra función de pérdida al aprender de los datos.

Resumen: función de pérdida

Una función de pérdida cuantifica qué tan insatisfecho estarías si usaras $f(x)$ para predecir Y^* cuando la salida correcta es Y . Es el objeto que queremos minimizar.



Regresión
Lineal

¿Suena familiar? Acabamos de repasar un proceso de optimización similar para la regresión lineal (**nuestra función de pérdida era MSE**).

Recuerda: Todos los modelos supervisados tienen una función de pérdida (a veces también conocida como función de costo) que debe ser optimizada cambiando los parámetros del modelo para minimizar esta función.

Metodología
de
aprendizaje

¿Cómo
aprende
nuestro árbol
de decisiones?

Los parámetros de un modelo son valores que nuestro modelo controla para minimizar nuestra función de pérdida. Nuestro modelo establece estos parámetros aprendiendo de los datos.



¿Cuánto \$\$ tengo?

< \$50

Cocinar en
casa

>= \$50

Comer en
restaurantes



Uno de los parámetros que controla nuestro modelo son los valores divididos.

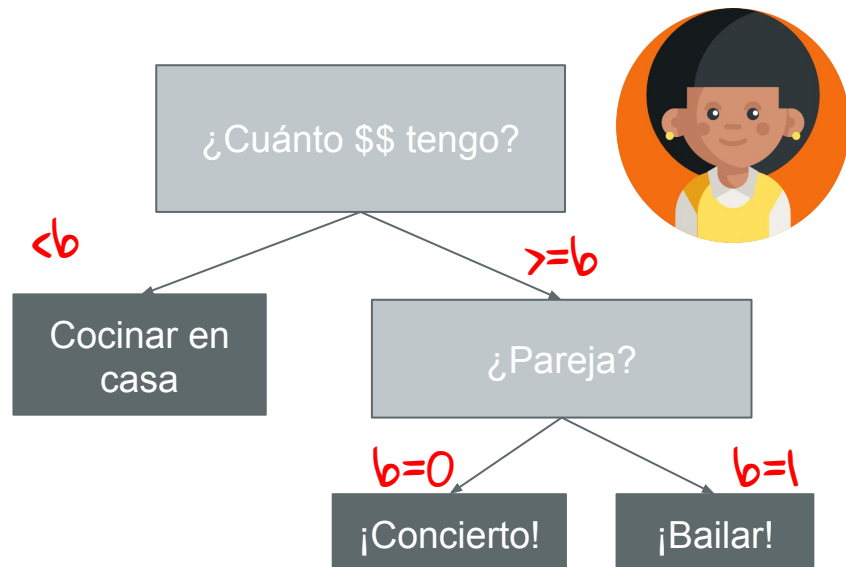
Por ejemplo, nuestro modelo aprende que \$ 50 es la mejor división para predecir si Sam come fuera o se queda en casa.

Parámetros

Valores que controlan el comportamiento del modelo. El modelo aprende cuáles son los parámetros de los datos.



Tenemos dos parámetros de árbol de decisión: regla de decisión dividida (valor de b) y el orden de las características



Nuestro modelo controla la regla de decisión de la división y el orden de las funciones.

Nuestro modelo aprende qué se ajusta mejor a los datos moviendo el valor dividido hacia arriba o hacia abajo y probando muchas combinaciones de ordenamiento de reglas de decisión.

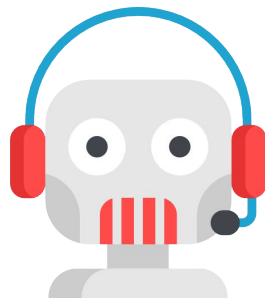
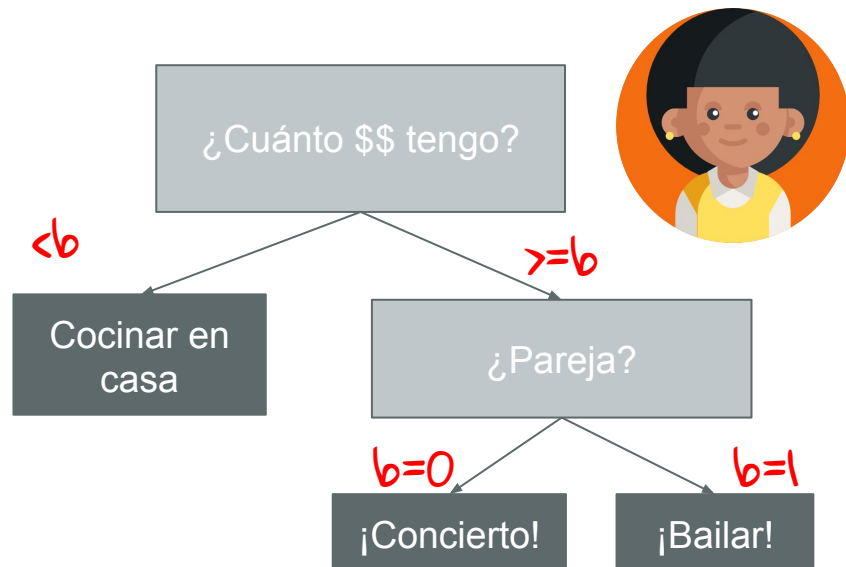
Problema central: ¿Cómo
aprendemos la “*mejor*” división?



Metodología
de
aprendizaje

¿Cómo
aprende
nuestro árbol
de decisiones?

Hay tantas rutas de decisión posibles y valores a dividir, de hecho, ¡una cantidad infinita! ¿Cómo elige nuestro modelo?



¡Oh no! Eso lo
empeoró. Probemos
otra cosa.

Nuestro modelo verifica las opciones de parámetros contra la función de pérdida cada vez que realiza un cambio.

Comprueba si el error subió, bajó o permaneció igual. Si cae, nuestro modelo sabe que hizo algo bien.

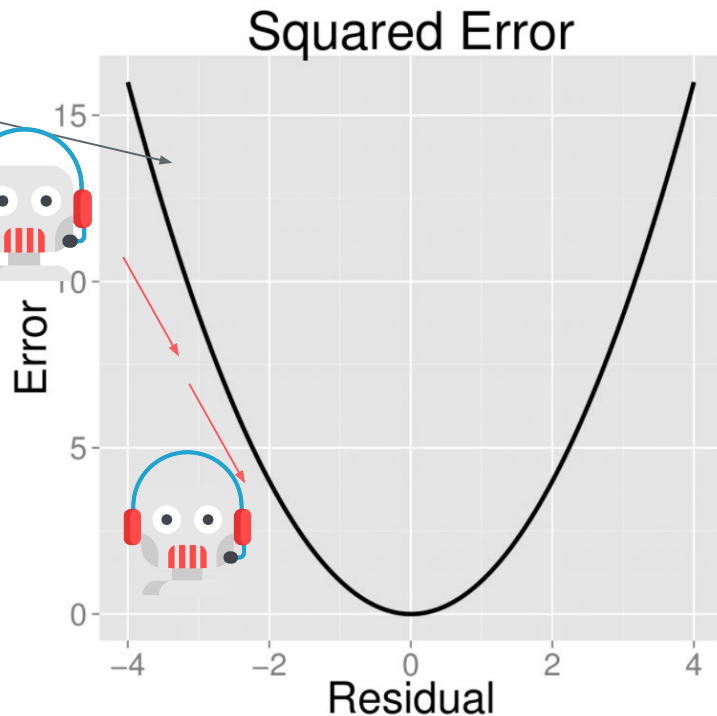


Imagina que es un juego. Comenzamos con un orden **aleatorio** de características y establecemos b para completar valores aleatorios.

Nuestra inicialización **aleatoria** de parámetros nos da un error inicial sorprendentemente alto

El trabajo de nuestro modelo es cambiar los parámetros para que cada vez que se actualice, la pérdida disminuya.

El juego termina cuando nuestro modelo puede reducir el error a e , el error irreducible.



Resumen de tareas: Los árboles de decisión también se pueden usar para dos tipos de tareas!



Regresión

Variable continua

Un problema de regresión es cuando estamos tratando de predecir un valor numérico dado alguna entrada, como “dinero” o “peso”.

Clasificación

Variable categórica

Un problema de clasificación es cuando se trata de predecir si algo pertenece a una categoría, como “rojo” o “azul” o “enfermedad” y “no enfermedad”.

Nuestra función de pérdida del árbol de decisión depende del tipo de tarea.

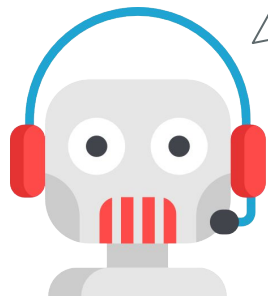
Las funciones de pérdida más comunes para los árboles de decisión incluyen:

Para árboles de clasificación:

1. Impureza de Gini
2. **Entropía**

Para árboles de regresión:

1. Error cuadrático medio

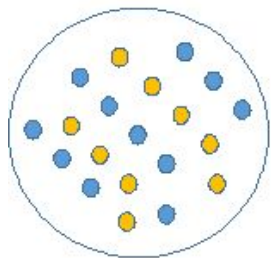


¡Puedo minimizar
todo tipo de
funciones de pérdida!

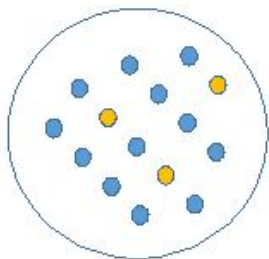
Por ejemplo, la función de pérdida para un árbol de decisión de regresión debería sentirse familiar. ¡Es la misma función de pérdida que usamos para la regresión lineal!

¡Para los árboles de clasificación, podemos usar la ganancia de información!

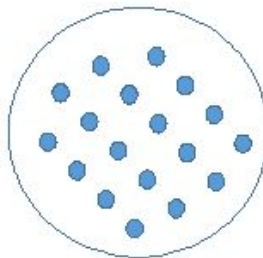
¿Cómo calificarías la entropía de estos círculos?



A



B



C

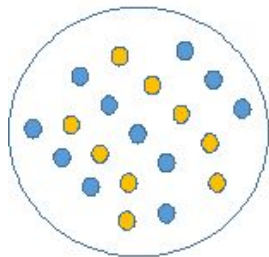
La ganancia de información intenta minimizar la **entropía** en cada subnodo. La entropía se define como un grado de desorganización.

Si la muestra es completamente homogénea, entonces la entropía es cero.

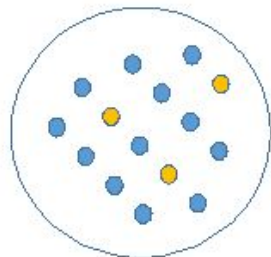
Si la muestra está dividida en partes iguales (50% - 50%), tiene una entropía de uno.

La teoría de la información cuantifica la cantidad de organización en términos de decisiones binarias o reglas necesarias para describir alguna estructura o patrón.

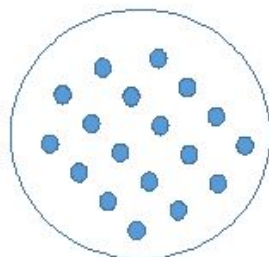
¿Cómo clasificarías los círculos en términos de entropía?



A



B



C

Entropía Alta

Entropía Mediana

Entropía Baja

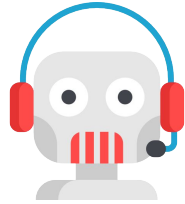
Cuanto más compleja o arbitraria es una estructura, más información se necesita para describirla.

Todas las cosas que son iguales deben guardarse en el mismo nodo terminal.

Nuestra población de baja entropía es completamente homogénea: ¡la entropía es 0!



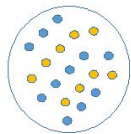
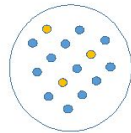
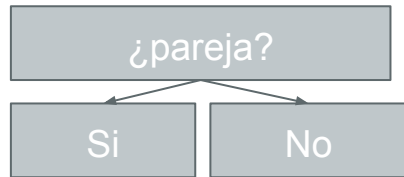
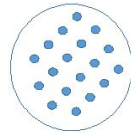
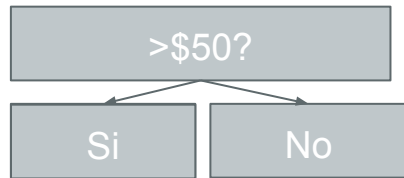
Entonces, ¿cómo funciona en nuestro modelo de árbol de decisión?



Voy a generar
aleatoriamente tres
divisiones ...

división

entropía



El algoritmo de **ganancia de información**:

1. Calcula la entropía del nodo primario
2. Calcula la entropía de cada nodo individual de división y calcula el promedio ponderado de todos los subnodos disponibles en división.

El algoritmo luego elige la división que tiene la entropía más baja.

Los árboles de regresión usan el error cuadrático medio.

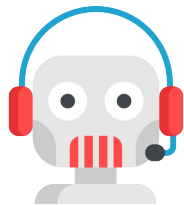
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$Y - Y^*$	For every point in our dataset, measure the difference between true Y and predicted Y.
2	Square each $Y - Y^*$ to get the absolute distance, so positive values don't cancel out negative ones when we sum.
Sum	Sum across all observations so we get the total error.
mean	Divide the sum by the number of observations we have.

This may look familiar - look back at our discussion of linear regression! Decision trees also use MSE.

The split with lower MSE is selected as the criteria to split the population.

Los árboles de regresión usan el error cuadrático medio.



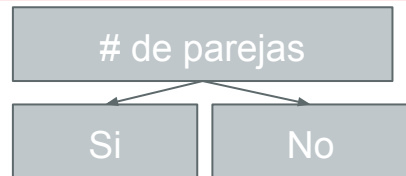
Voy a generar
aleatoriamente tres
divisiones ...

División

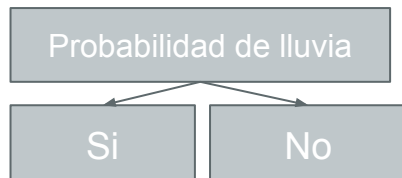
RMSE



3.83



7.28



10.3

El algoritmo RMSE procede de la siguiente manera:

1. Cálculo de la varianza para cada nodo.
2. Cálculo de la varianza para cada división como **promedio ponderado** de cada varianza de nodo.

El algoritmo luego selecciona la división que tiene la varianza más baja.

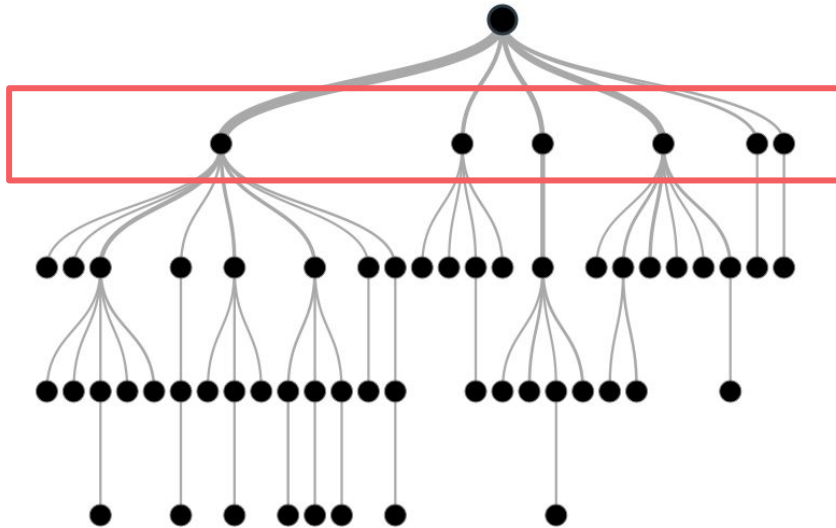
Desempeño del modelo



Desempeño

Desempeño de las
características

Los árboles de decisión nos permiten comprender qué características son más importantes para predecir Y *



La intuición es proporcionada por la comprensión de que las divisiones más importantes se encuentran en los primeros nodos.

Recordemos nuestra intuición: las divisiones importantes suceden más cerca del nodo raíz.

El árbol de decisión nos dice cuáles son las divisiones importantes (qué características causan divisiones que reducen **más** la función de costos).

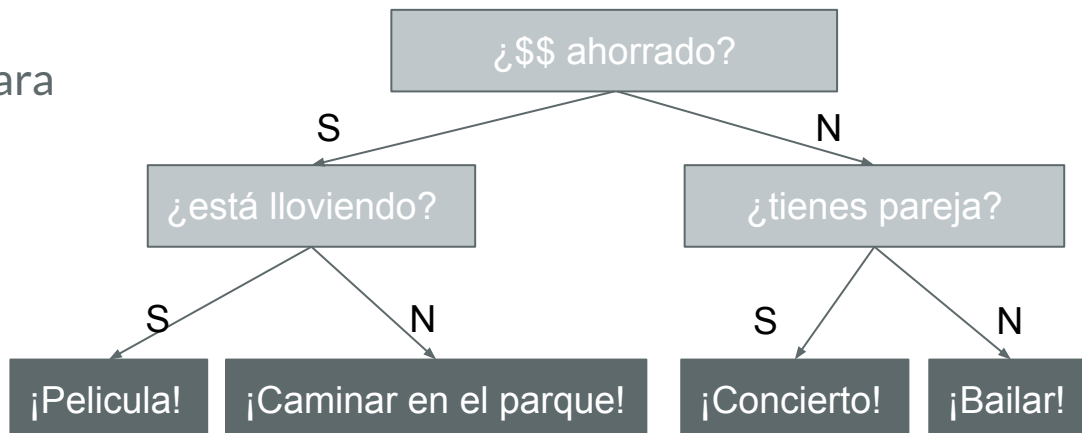
Desempeño

Desempeño de las
características

Recordemos nuestra intuición: las
divisiones importantes suceden
antes

En el caso de Sam, nuestro
algoritmo determinó que el
presupuesto de Sam era la
característica más importante para
predecir sus planes de fin de
semana.

¡Esto tiene mucho sentido!

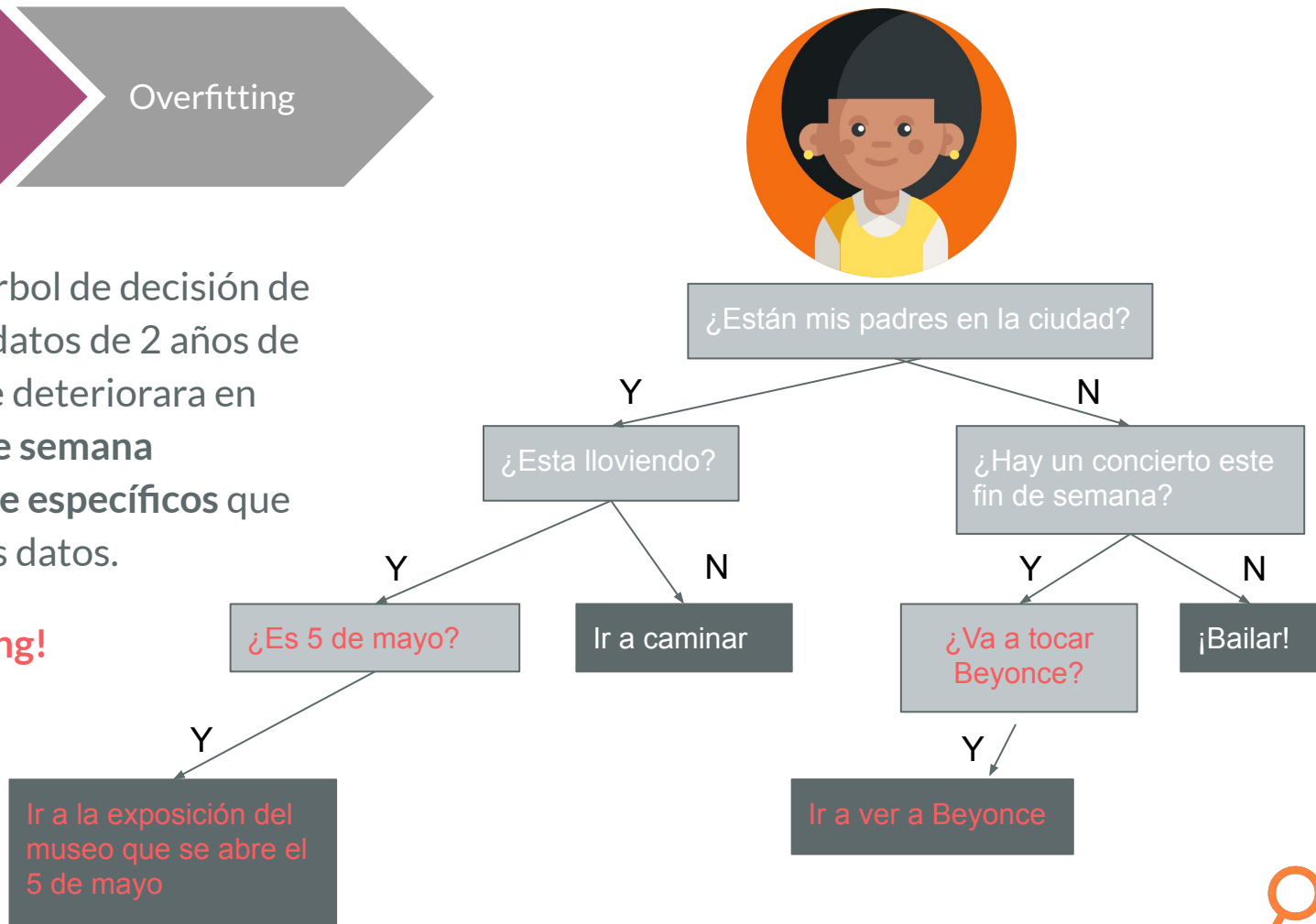


Desempeño

Overfitting

Imagínesse si el árbol de decisión de Sam, basado en datos de 2 años de fin de semana, se deteriorara en **eventos de fin de semana extremadamente específicos** que ocurrieron en los datos.

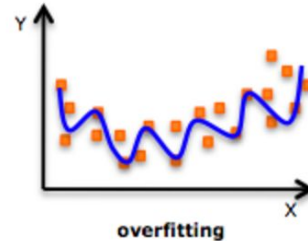
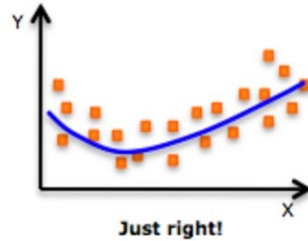
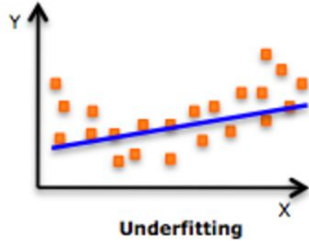
Esto es overfitting!



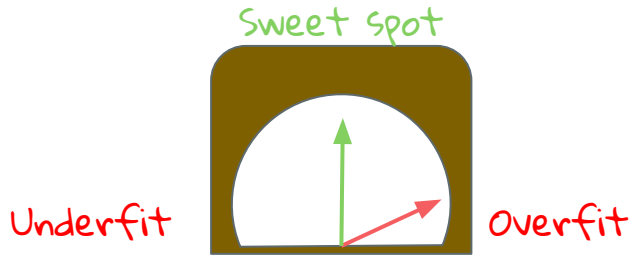
Desempeño

Capacidad de
generalizar a
datos no vistos

Recordar que el objetivo más importante que tenemos es construir un modelo que se generalice bien a datos invisibles.



Si nuestro conjunto de datos de entrenamiento se sobreajusta, no se generalizará bien a nuestro conjunto de prueba (datos no vistos).



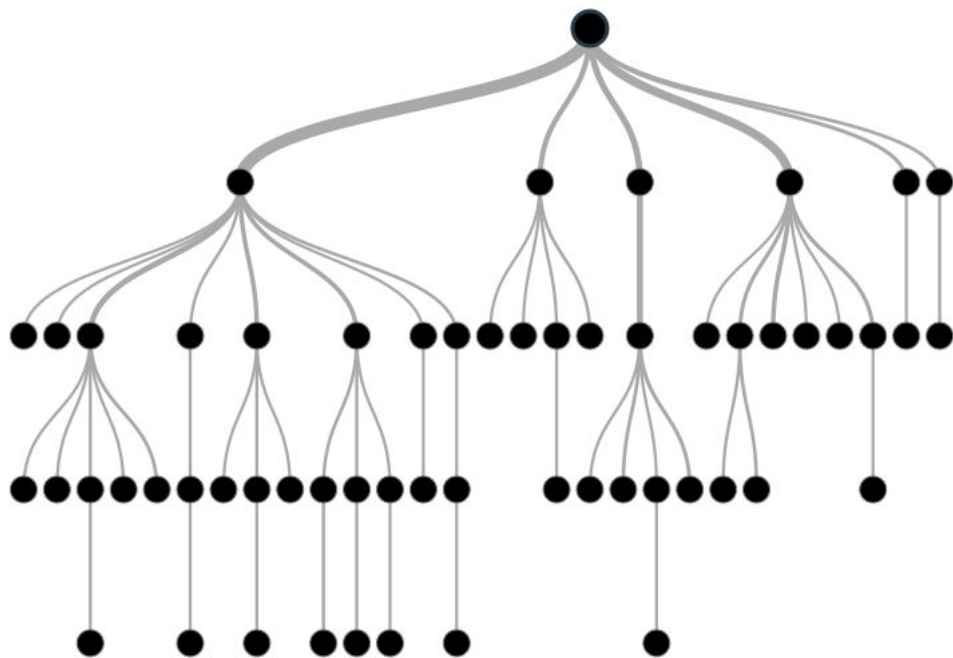
Nuestro objetivo en la evaluación del rendimiento es encontrar un punto óptimo entre el sobreajuste y el subajuste.

Sin embargo, si se subajusta, no estamos capturando la complejidad de la relación y nuestra pérdida será alta.

Desempeño

Capacidad de
generalizar a
datos no vistos

Desafortunadamente, los árboles de decisión son muy propensos al **sobreajuste**. Tenemos que tener mucho cuidado al evaluar esto.



Cada vez que agregamos un nodo, ajustamos modelos adicionales a subconjuntos de datos. Esto significa que comenzamos a conocer muy bien los datos de nuestro entrenamiento, pero afectará nuestra capacidad de generalizar a nuestros datos de prueba.

Sobreajuste: si cada nodo terminal es una observación individual, está sobreajustado.

¡Veamos un ejemplo concreto de optimización de nuestro modelo!



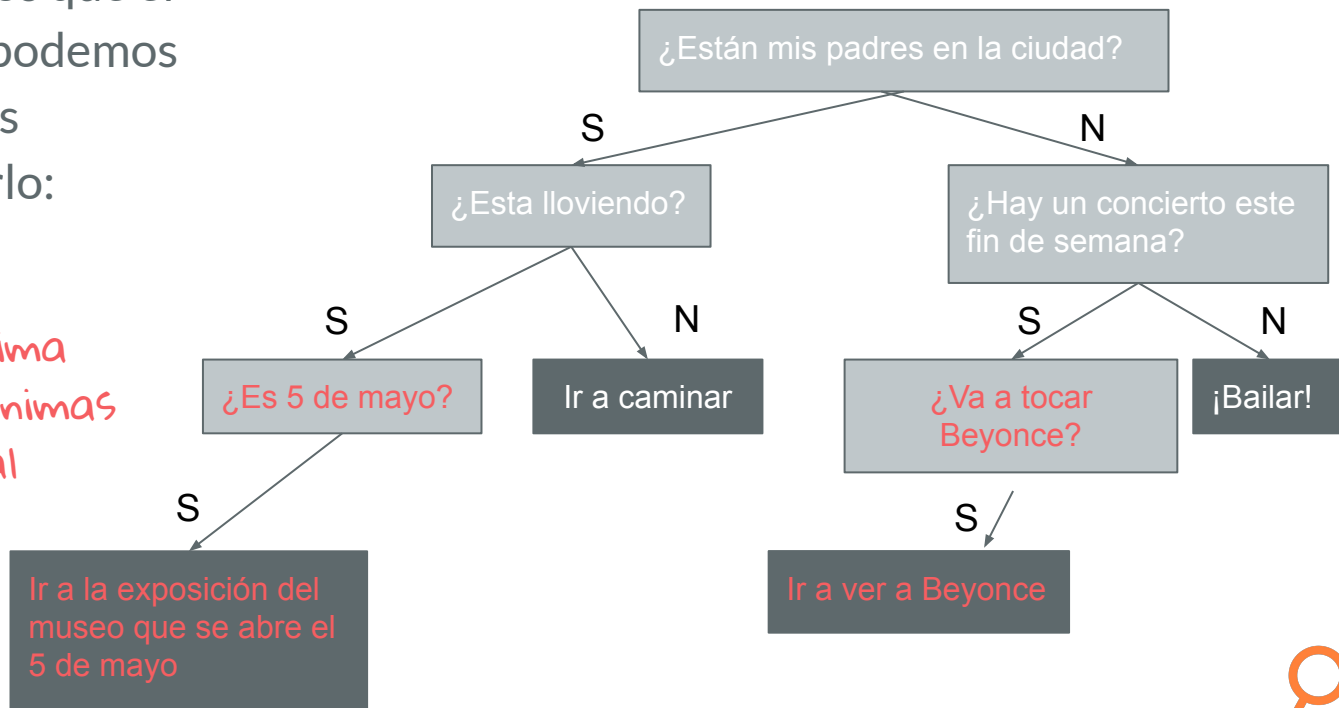
Optimización de modelo





Ahora que entendemos que el sobreajuste es malo, podemos usar algunos enfoques diferentes para evitarlo:

- Poda
- Profundidad máxima
- Observaciones mínimas por nodo terminal



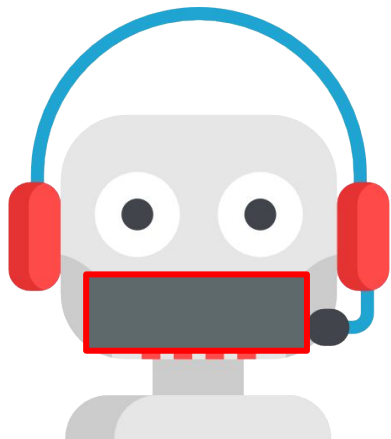
Poda, profundidad máxima y n_obs en un nodo terminal son ejemplos de hiperparámetros.

Los hiperparámetros son configuraciones de nivel superior de un modelo que se arreglan antes de que comience el entrenamiento.

Poda, profundidad máxima y n_obs en un nodo terminal son hiperparámetros del árbol de decisión establecidos antes del entrenamiento.

Sus valores no se aprenden de los datos, por lo que el modelo no puede decir nada sobre ellos.

¡TÚ, no el modelo,
decide cuáles son los
hiperparámetros!



Resumen: ¿Qué es un hiperparámetro?

- En regresión lineal, los coeficientes eran **parámetros**.

$$Y = a + b_1x_1 + b_2x_2 + \dots + e$$

El modelo decide!

- Los parámetros se **aprenden** de los datos de entrenamiento utilizando el algoritmo elegido.
- Los hiperparámetros no se pueden aprender del proceso de capacitación. Expresan propiedades de “nivel-superior” del modelo, como su **complejidad** o la **rapidez con que debe aprender**.

Nosotros decidimos!

Metodología
de
aprendizaje

Hiperparámetros
del árbol de
decisiones

Poda

De arriba hacia
abajo

De abajo hacia
arriba

Máxima profundidad

Muestras mínimas
por hoja

Número máximo de
nodos terminales

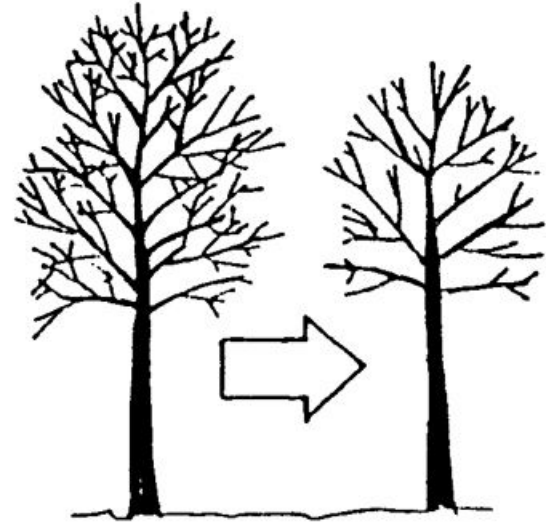


La poda reduce el número de divisiones en el árbol de decisión.

Limitar el número de divisiones minimiza el problema del **sobreajuste (overfitting)**.

Dos enfoques:

- Pre-pruning (de arriba hacia abajo)
- Post-pruning (de abajo hacia arriba)



Lee más: [Notre Dame's Data Mining class, CSE 40647](#)

La pre-poda es un enfoque de arriba hacia abajo donde el árbol tiene una profundidad limitada antes de que crezca por completo.

Pre-poda ralentiza el algoritmo antes de que se convierta en un árbol completamente desarrollado. Esto evita divisiones irrelevantes.

- P.ej. En el ejemplo del diagnóstico de malaria, probablemente no tendría sentido incluir preguntas sobre el color favorito de una persona.
 - Podría causar una división, pero probablemente no sería una división significativa. Puede conducir a un sobreajuste

Una analogía útil en la naturaleza es un árbol bonsai. Este es un árbol cuyo crecimiento se ralentiza a partir de cuando es un árbol joven.



Hiperparámetros

post-Poda

Post-poda es un enfoque ascendente en el que el árbol tiene una profundidad limitada después de crecer completamente.

Cultivar el árbol completo, luego podar fusionando nodos terminales.

1. Dividir datos en conjunto de entrenamiento y validación
2. Usando el árbol de decisión obtenido del conjunto de entrenamiento, combine dos nodos terminales
3. Calcular el error del árbol con nodos fusionados y árbol sin nodos fusionados
4. Si el árbol con nodos combinados tiene un error menor, combinar los nodos hoja y repetir 2-4.

Una analogía útil en la naturaleza son los arbustos podados. Los arbustos crecen a su máximo potencial, luego se cortan y se les da forma.

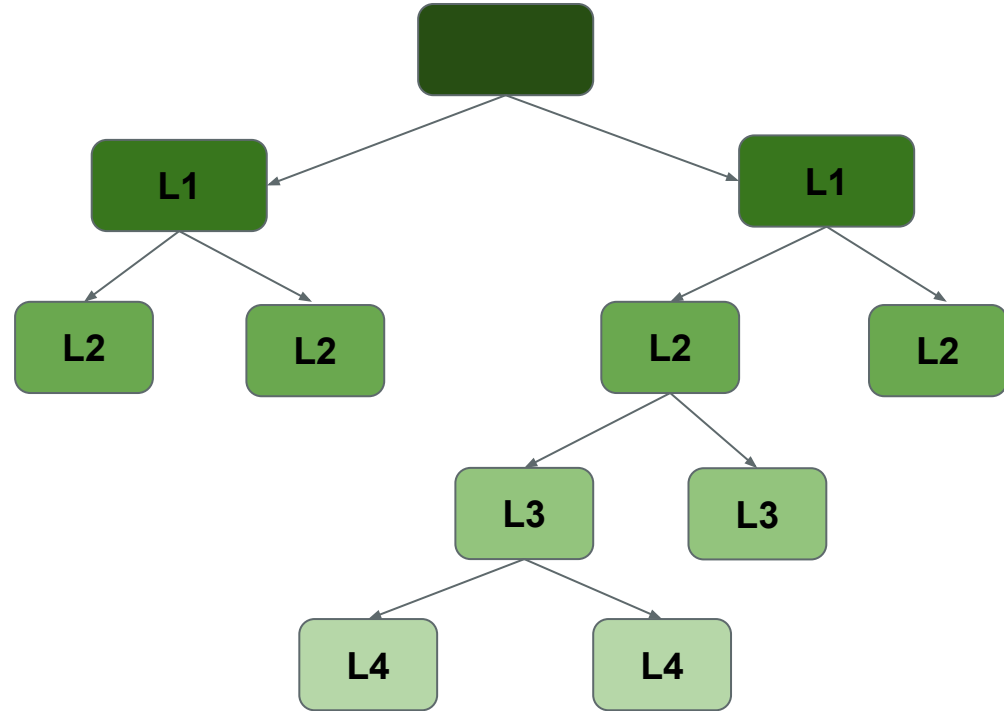


Hiperparámetros

Máxima
profundidad

La profundidad máxima define el número máximo de capas que puede tener un árbol.

Por ejemplo, una profundidad máxima de 4 significa que un árbol se puede dividir entre 1 y 4 veces.

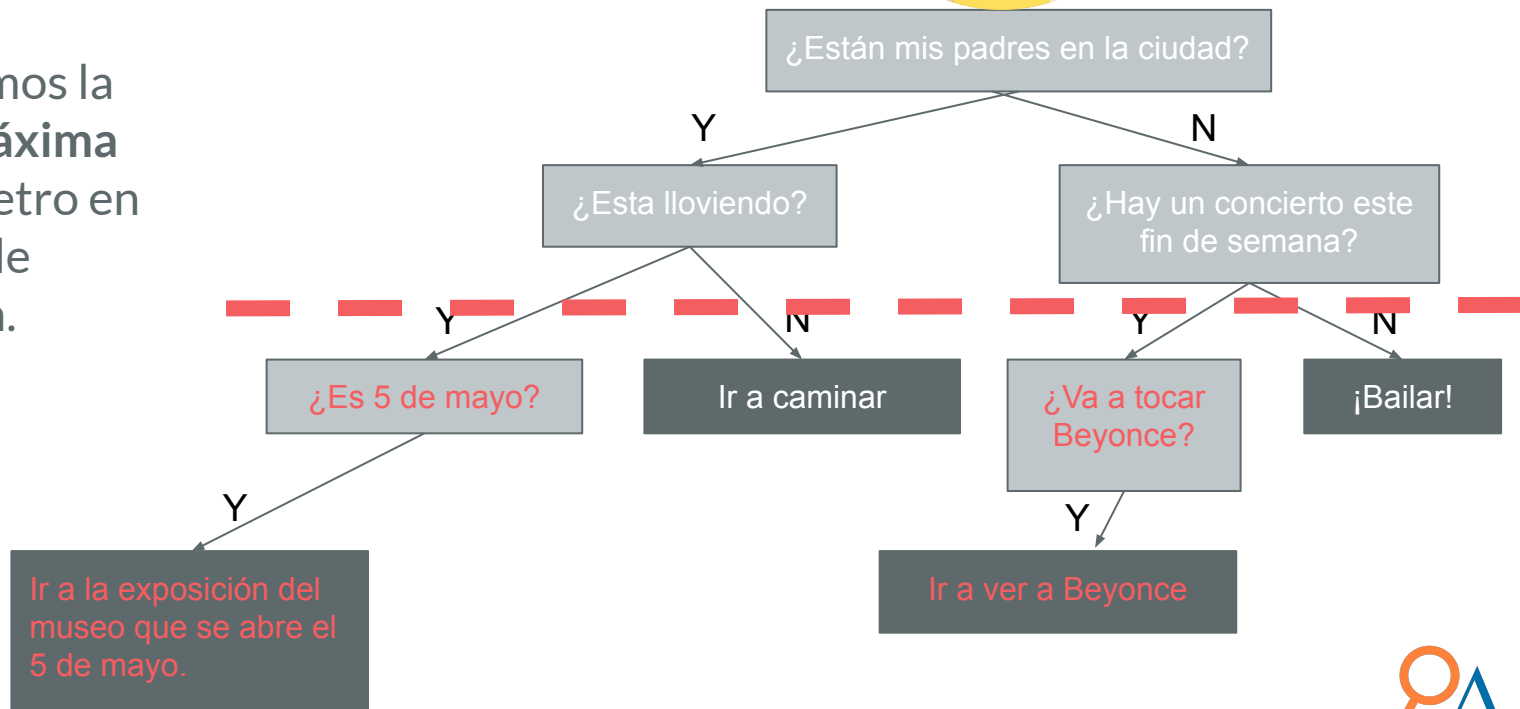


Profundidad máxima alcanzada, no más división



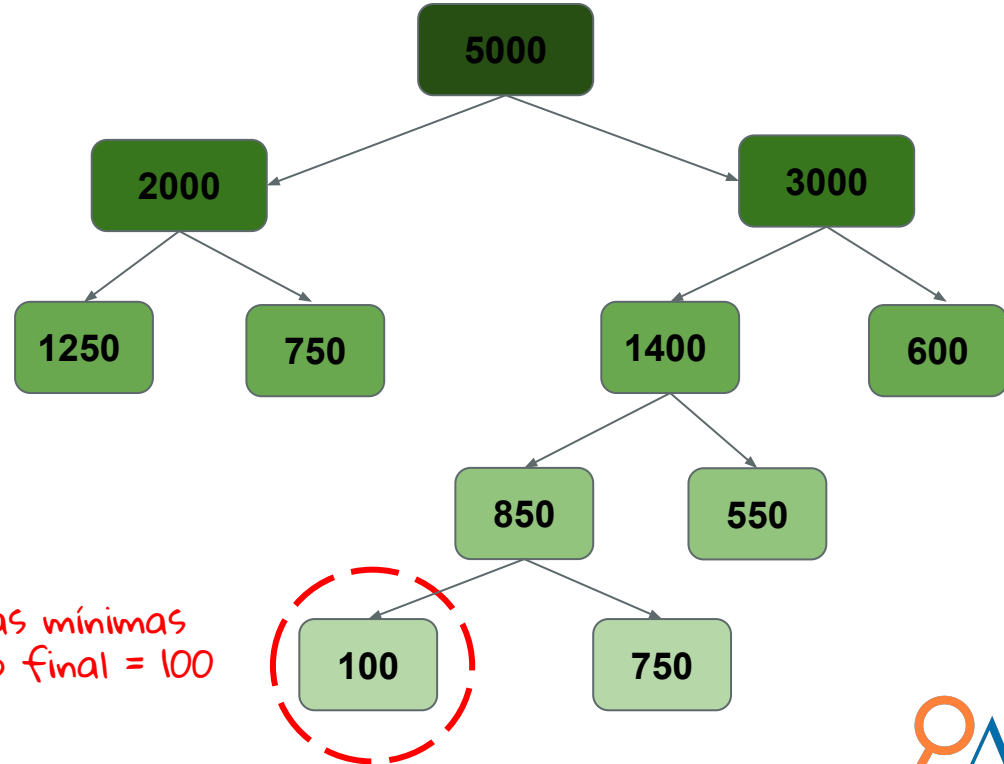


Aquí establecemos la **profundidad máxima** del hiperparámetro en 2 para el árbol de decisión de Sam.



Observaciones mínimas por nodo final

Este hiperparámetro establece un número mínimo de observaciones en un nodo final, lo que reduce la posibilidad de modelar el ruido de los datos.



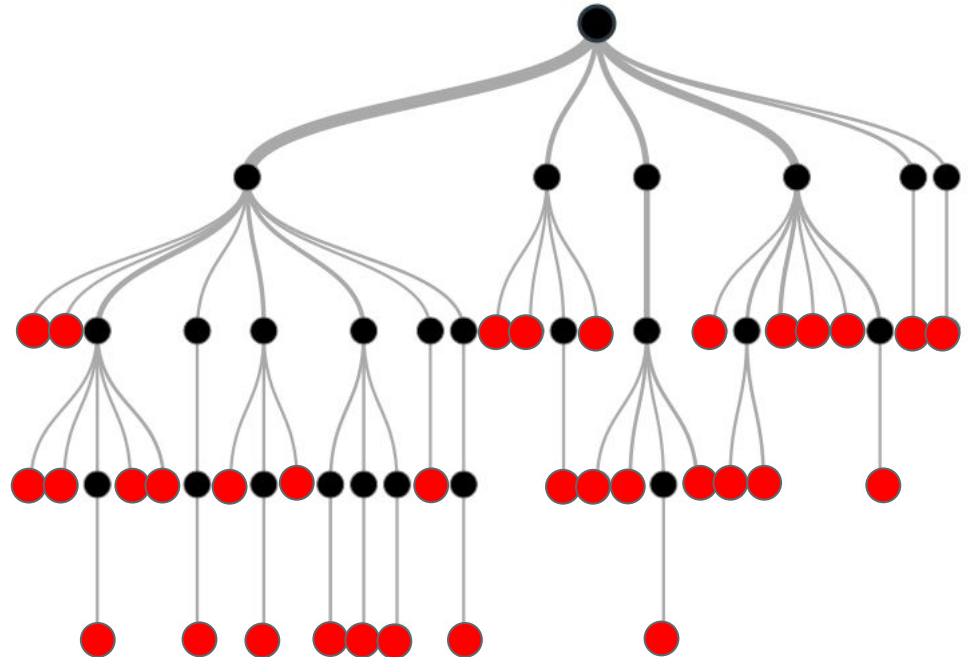
Hiperparámetros

Número
máximo de
nodos
terminales

El número máximo de nodos
terminales limita el número de
ramas en nuestro árbol.

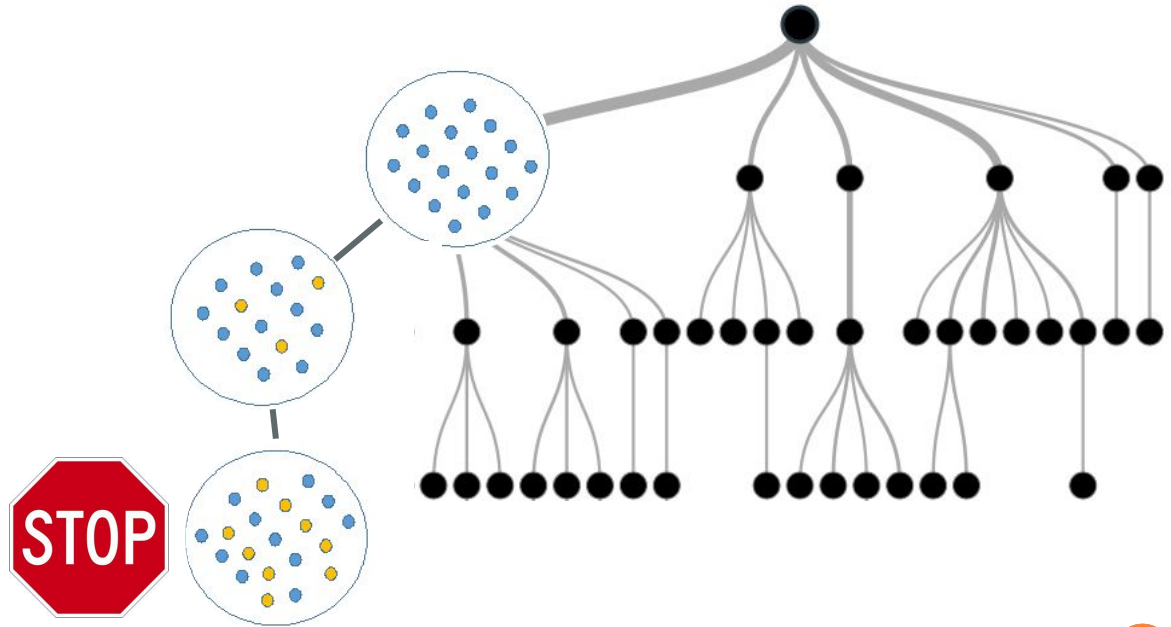
Nuevamente, esto limita
el sobreajuste y reduce la
posibilidad de ruido de
modelado.

● = nodos terminales



Impureza mínima

- Recordar la función de costo de ganancia de información
- El modelo itera hasta que alcanza un cierto nivel de impureza.



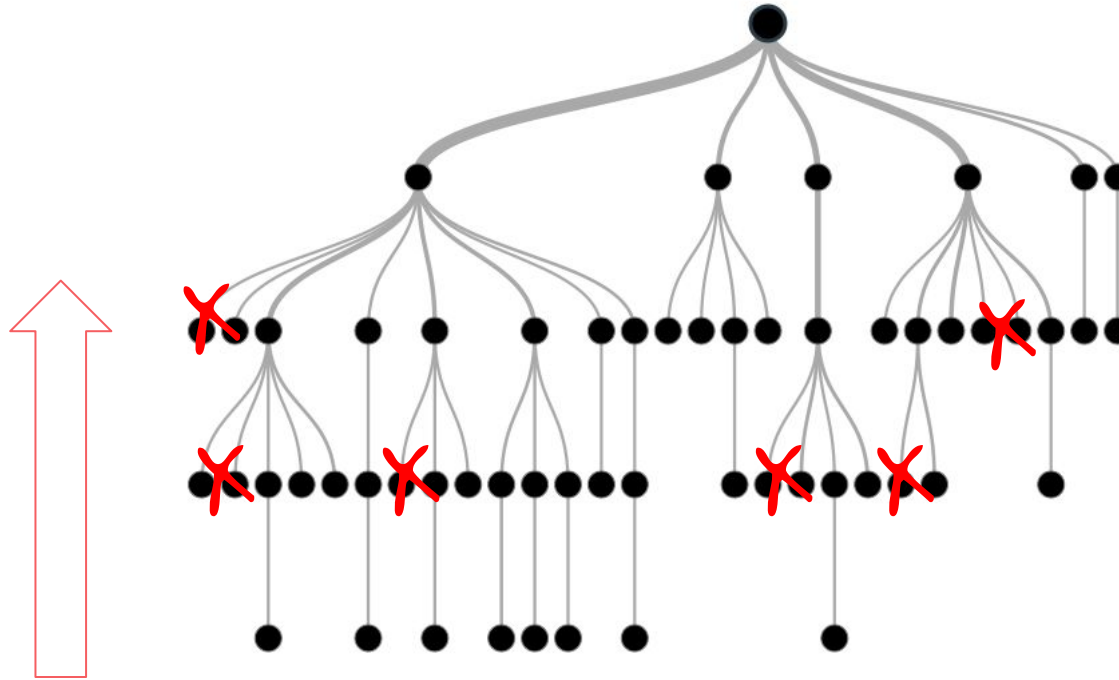
¿Qué hiperparámetro debo usar?

No hay hiperparámetro objetivamente mejor para cada modelo.
Use prueba y error - ¡ve cómo cada uno cambia sus resultados!

Metodología
de
aprendizaje

Proceso de
optimización

Poda de abajo hacia



Combina selectivamente algunos nodos terminales



Poda de abajo hacia

Veamos cómo funciona una sola fusión:

El modelo selecciona aleatoriamente la fusión marcada en el árbol.

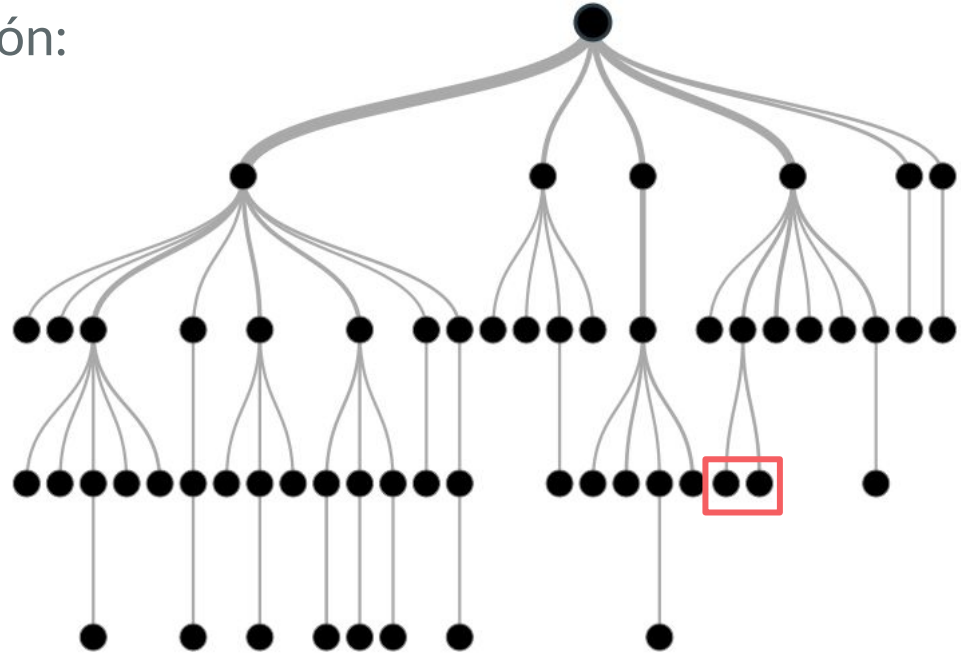
El modelo calcula:

Error de árbol sin fusión = a

Error de árbol con fusión = b

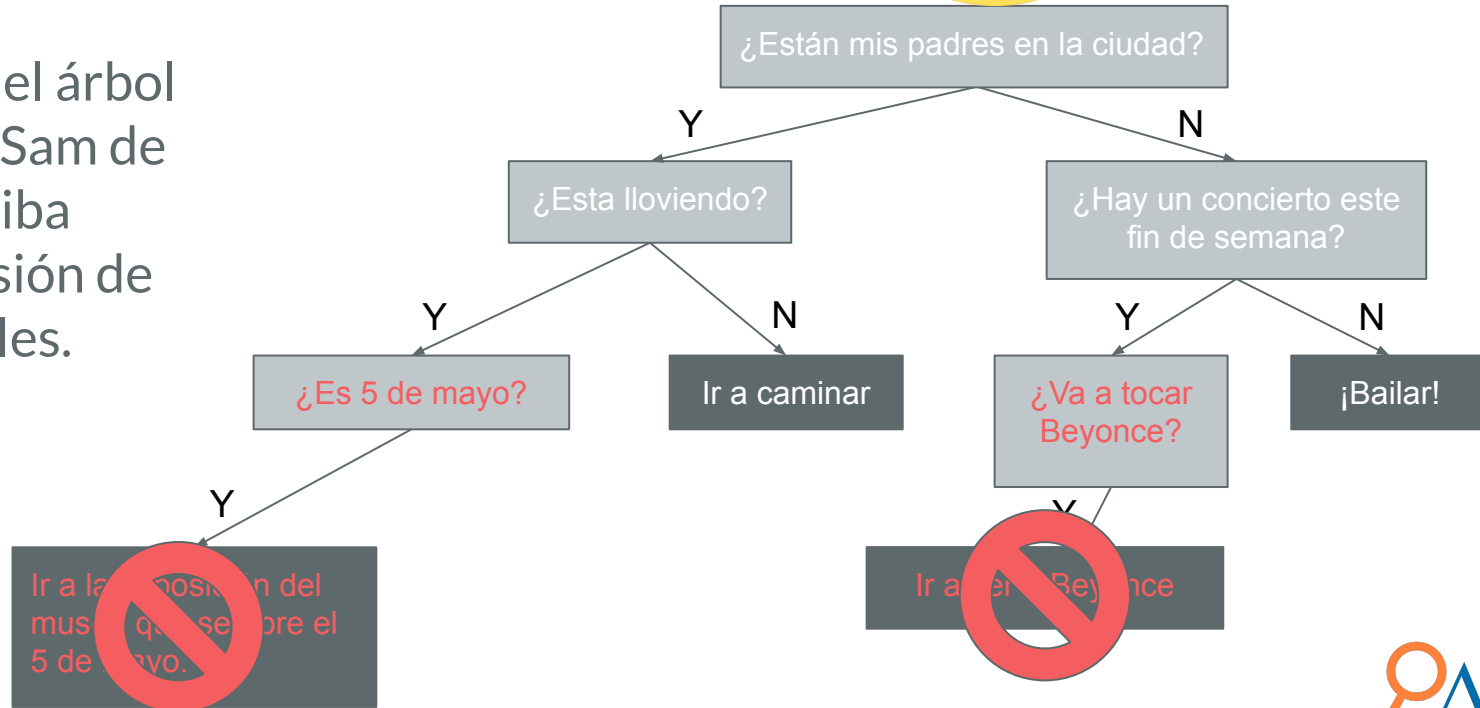
Si $b < a$, ¡elige fusionarse!

... Y repetir

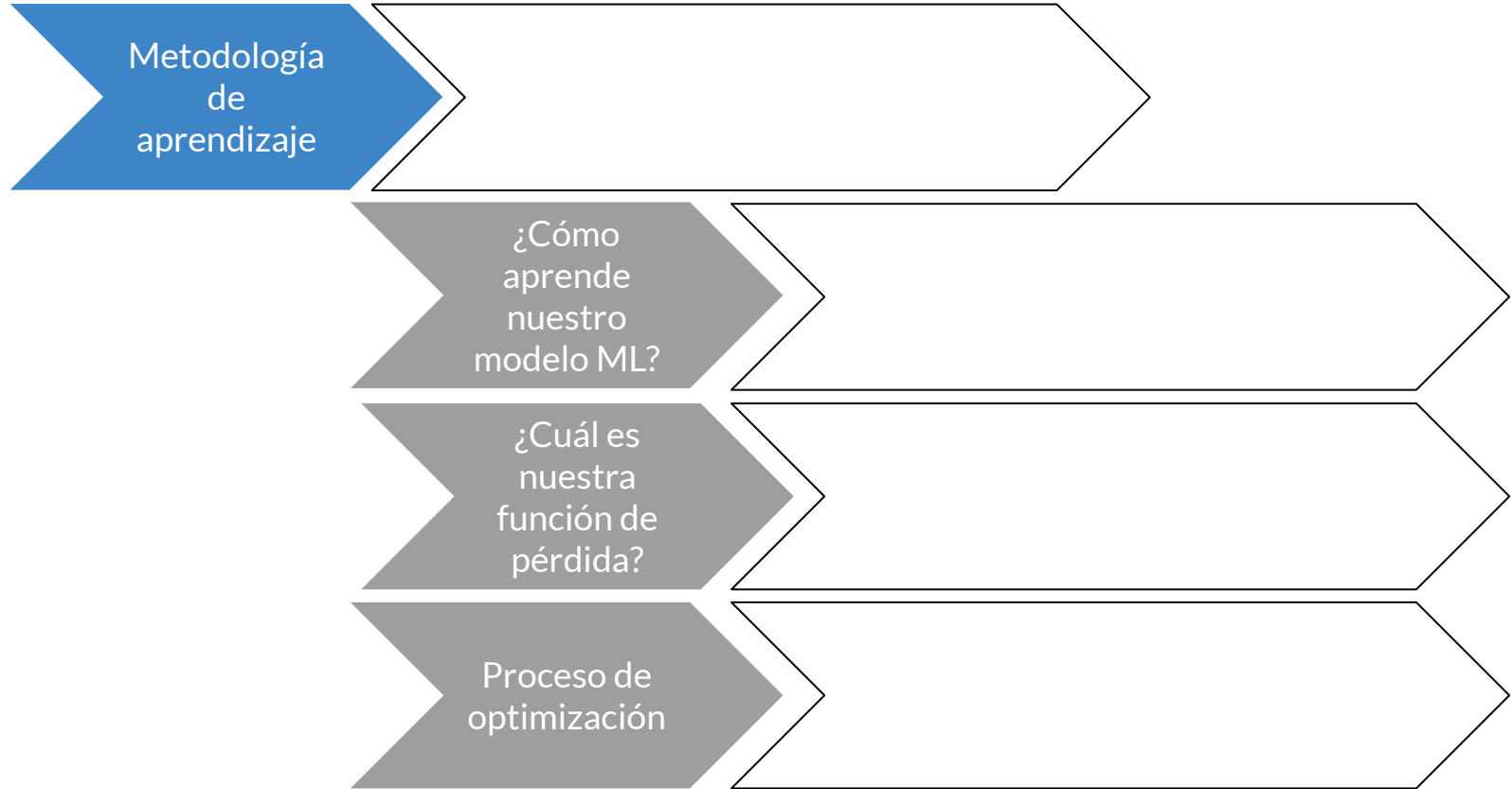




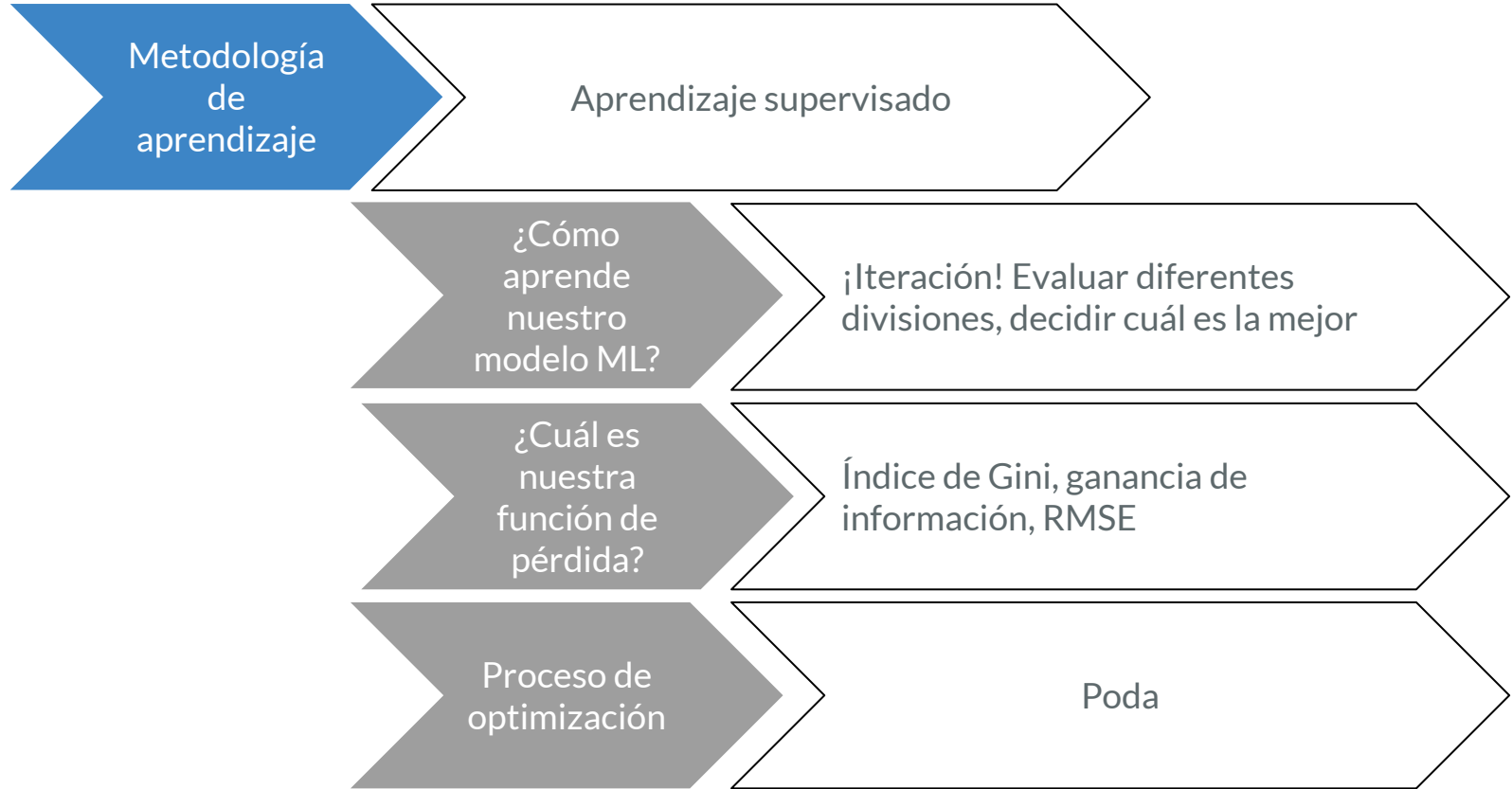
Aquí podemos el árbol de decisión de Sam de abajo hacia arriba mediante la fusión de nodos terminales.



Learning Methodology



Learning Methodology



Fin del módulo.



Checklist:

- ✓ Árboles de decisión
 - ✓ Intuición
 - ✓ La “mejor” división
 - ✓ Desempeño del modelo
 - ✓ Optimización del modelo (poda)



Recursos Avanzados



¿Quieres ir más allá? Aquí hay algunos recursos que recomendamos:

- Libros

- [An Introduction to Statistical Learning with Applications in R \(James, Witten, Hastie and Tibshirani\)](#): Chapter 8
- [The Elements of Statistical Learning: Data Mining, Inference, and Prediction \(Hastie, Tibshirani, Friedman\)](#): Chapter 9

- Recursos en línea

- [Analytics Vidhya's guide to understanding tree-based methods](#)



Felicidades! ¡Terminaste el módulo!

Obtén más información sobre el machine learning de Delta para una buena misión aquí.