



Pontificia Universidad
JAVERIANA
Bogotá

Técnicas Predictivas para la Generación de Nuevo Conocimiento

ING. HERIBERTO FELIZZOLA JIMENEZ

Educación **Continua**
Generamos experiencias educativas

Contenido

1. Introducción al Aprendizaje de Máquina
2. Aprendizaje Supervisado (Técnicas Predictivas)
3. Modelo de Regresión Lineal
4. Arboles de Clasificación
5. Métodos de Ensamble



Pontificia Universidad
JAVERIANA
Bogotá

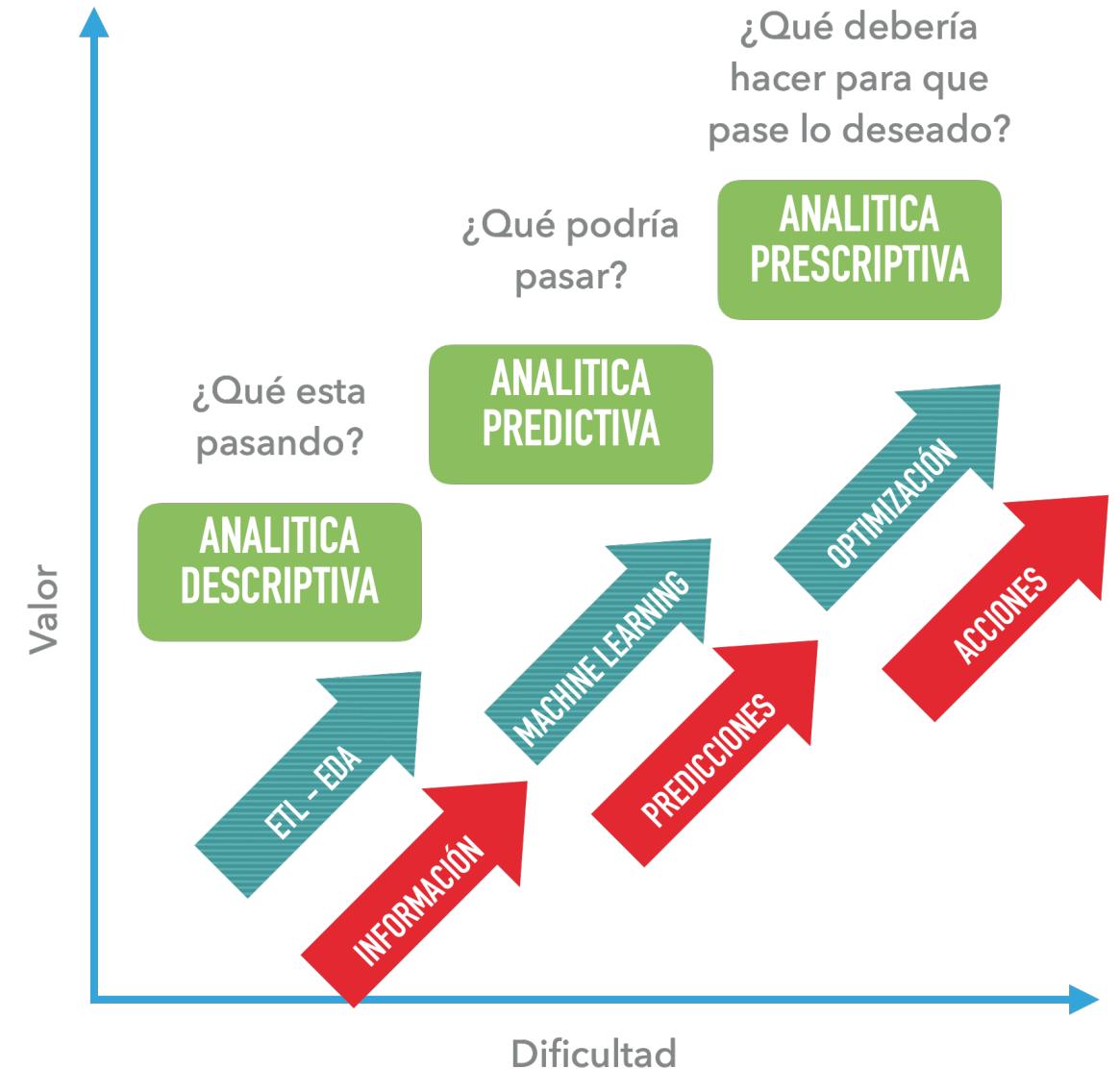
1. Introducción al Aprendizaje de Máquina

ING. HERIBERTO FELIZZOLA JIMENEZ

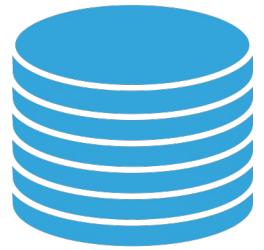
Educación **Continua**
Generamos experiencias educativas

Analítica Predictiva

- Es el arte de construir modelos que permiten realizar predicciones a partir de los patrones encontrados en los datos.
- Algunas aplicaciones de las predicciones son:
 - Precios de productos, servicios y mercado de valores
 - Recomendación de productos
 - Sistemas de recomendación
 - Riesgo en proyectos
 - Diagnóstico de enfermedades
 - Clasificación de imágenes, texto y video
 - Condiciones meteorológicas

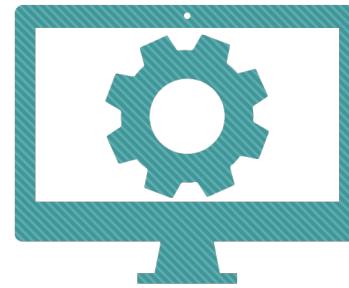


Aprendizaje de Máquina



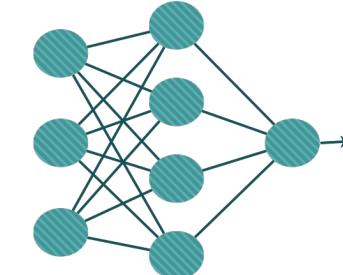
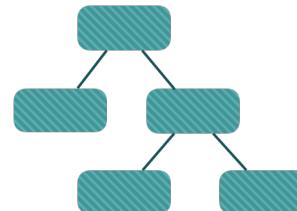
Datos de Entrenamiento

X ₁	X ₂	...	X _p	Y
X ₁₁	X ₁₂	...	X _{1p}	Y ₁
X ₂₁	X ₂₂	...	X _{2p}	Y ₃
...
X _{n1}	X _{n2}	...	X _{np}	Y _n

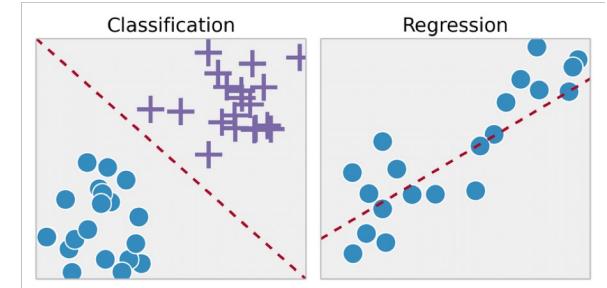


Algoritmos y Modelos de Aprendizaje de Maquina

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$



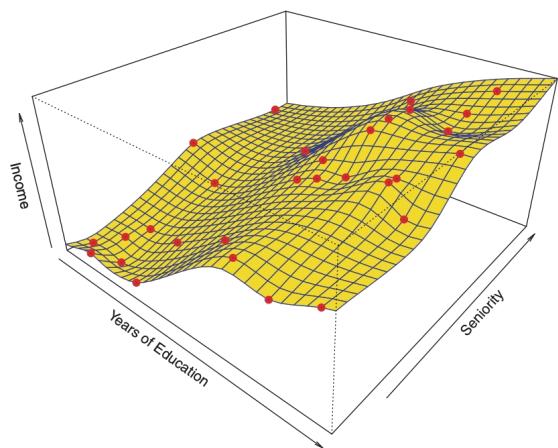
Predicciones



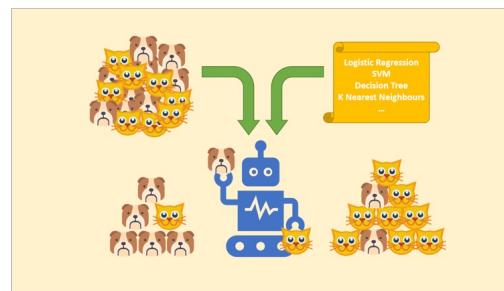
Tareas en Aprendizaje Automático

APRENDIZAJE SUPERVISADO

REGRESIÓN

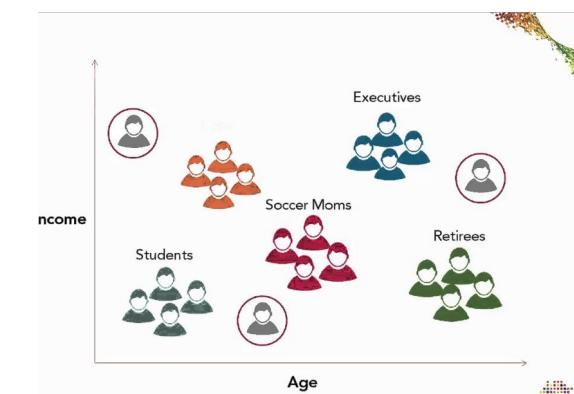


CLASIFICACIÓN

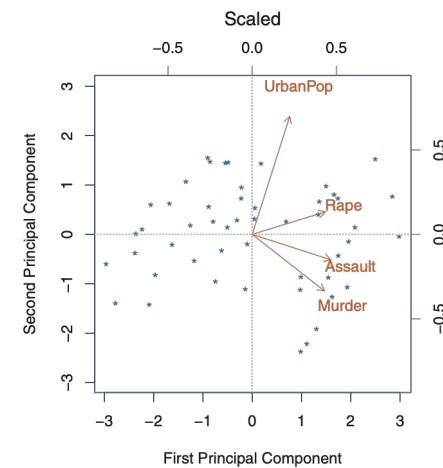


APRENDIZAJE NO SUPERVISADO

CLUSTERING



REDUCCIÓN DE DIMENSIONES

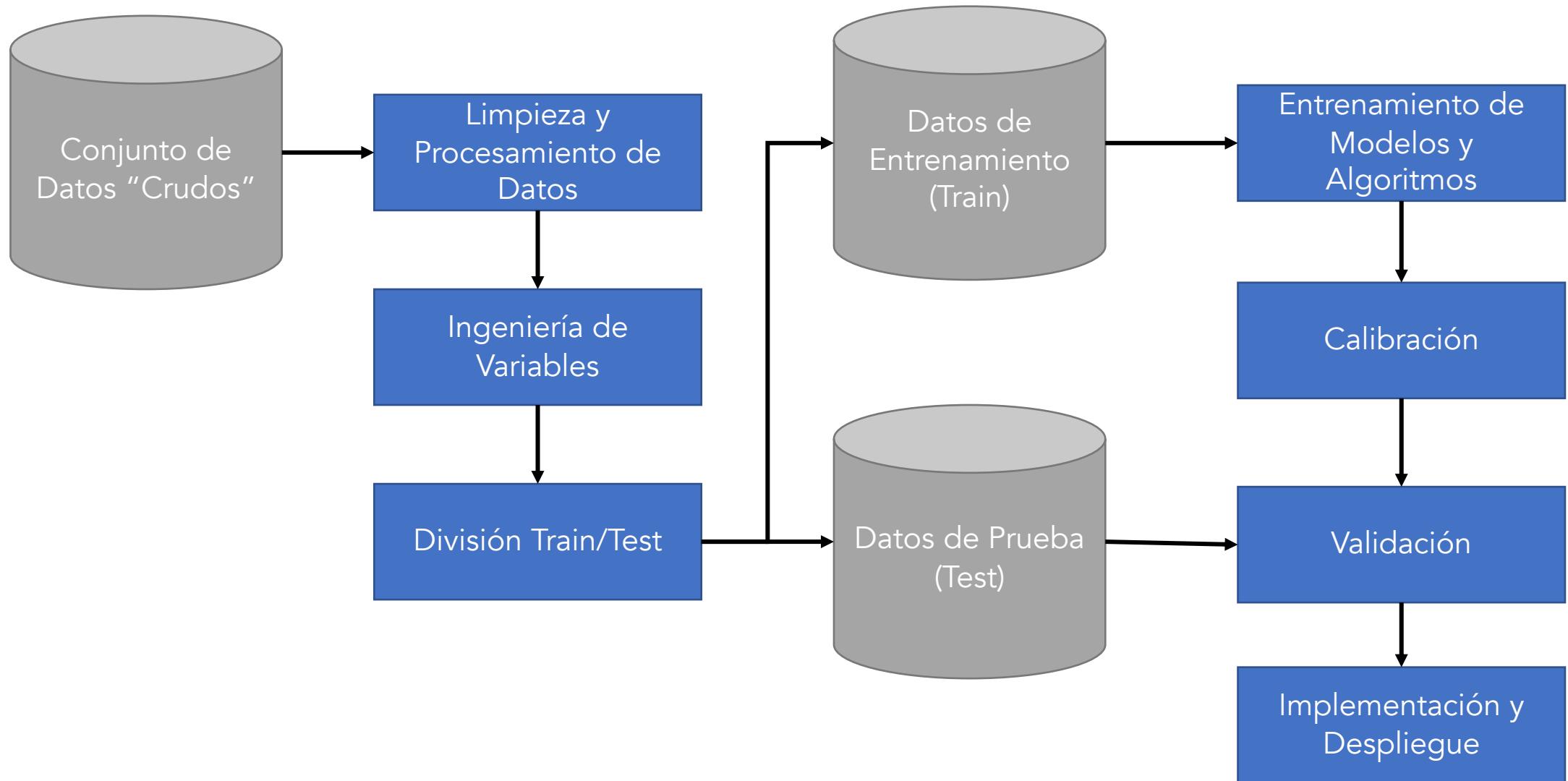


Instancia - Atributo - Salida

- La entrada a un esquema de aprendizaje automático es un conjunto de instancias.
- Estas instancias son las cosas que deben clasificarse, asociarse o agruparse.
- El conjunto de características de cada instancia es denominado atributos o variables de entrada (X).
- El resultado que se quiere predecir es denominado salida (Y).

Instancias	X: Atributos - Variables de Entrada				
	Outlook	Temperature	Humidity	Windy	Play Time
Sunny	85	85	false	5	
Sunny	80	90	true	0	
Overcast	83	86	false	55	
Rainy	70	96	false	40	
Rainy	68	80	false	65	
Rainy	65	70	true	45	
Overcast	64	65	true	60	
Sunny	72	95	false	0	
Sunny	69	70	false	70	
Rainy	75	80	false	45	
Sunny	75	70	true	50	
Overcast	72	90	true	55	
Overcast	81	75	false	75	
Rainy	71	91	true	10	

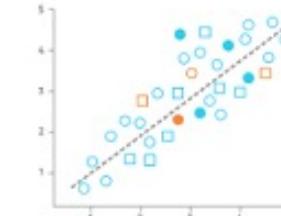
Proceso de Entrenamiento



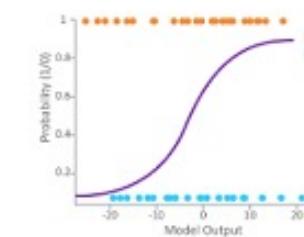
Modelos y Algoritmos

- Aprendizaje supervisado:
 - Modelos lineales: Mínimos Cuadrados, Análisis discriminante, Regresión Logística.
 - Modelos de penalización: Ridge, Lasso, ElasticNet
 - Modelos no lineales: Splines, Regresión Local, Modelos Generales Aditivos (GAM), Kernels.
 - Arboles: CART, C5.0, Bagging, Random Rorest, Boosting
 - Maquinas de soporte vectorial SVM.
 - Naive-Bayes
 - Redes neuronales.
- Aprendizaje no supervisado:
 - Algoritmos de clustering: K-mean, Modelos Jerarquicos
 - Reducción de dimensiones: Componentes principales - PCA

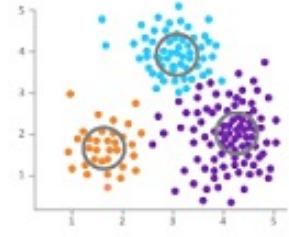
Linear Regression



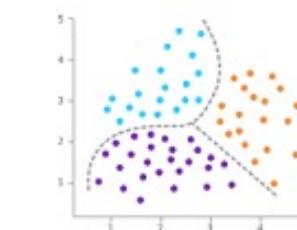
Logistic Regression



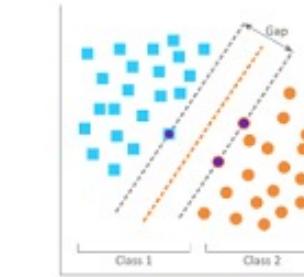
K-Means



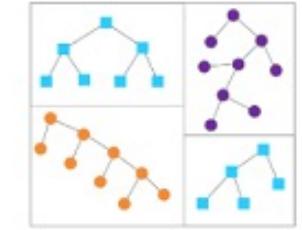
Naive Bayes



Support Vector Machines



Random Forest





Pontificia Universidad
JAVERIANA
Bogotá

2. Aprendizaje Supervisado Técnicas Predictivas

ING. HERIBERTO FELIZZOLA JIMENEZ

Educación **Continua**

Generamos experiencias educativas



Pontificia Universidad
JAVERIANA
Bogotá

2.1 Regresión Lineal

Educación **Continua**
Generamos experiencias educativas

Definición:

Un modelo de regresión es una técnica estadística utilizada para predecir el valor de una variable dependiente (objetivo) basado en uno o más valores de variables independientes (predictoras).

Propósito:

- Estimar relaciones entre variables.
- Predecir valores futuros basados en datos históricos.
- Identificar tendencias y patrones.

Tipos de Problemas que Resuelven:

- Problemas de predicción continua (p.ej., precios, temperaturas).
- Modelo de relaciones lineales y no lineales.
- Evaluación de la influencia de variables predictoras.

Algunas Aplicaciones



1. Predicción de Precios de Vivienda:

Descripción: Utilizar características de las propiedades (p.ej., tamaño, número de habitaciones, ubicación) para predecir el precio de venta.

Ejemplo: Estimar el precio de una casa en función de su tamaño en metros cuadrados, número de dormitorios y ubicación.



2. Nivel de Gastos o Compras de Clientes a Través de Productos Financieros:

Descripción: Predecir el comportamiento de gasto de los clientes basado en su historial de compras y características demográficas.

Ejemplo: Determinar el gasto mensual de un cliente utilizando su historial de transacciones y datos personales como edad e ingresos.



3. Predicción de Ventas:

Descripción: Estimar futuras ventas de un producto basado en datos históricos y factores de mercado.

Ejemplo: Predecir las ventas de un producto durante la temporada navideña en función de las ventas de años anteriores y el presupuesto publicitario actual.



4. Predicción de Costos Médicos:

Descripción: Estimar los costos futuros de atención médica basados en el historial médico y las condiciones de salud de los pacientes.

Ejemplo: Predecir los costos anuales de atención médica para un paciente con diabetes utilizando datos de su historial médico y tratamientos previos.



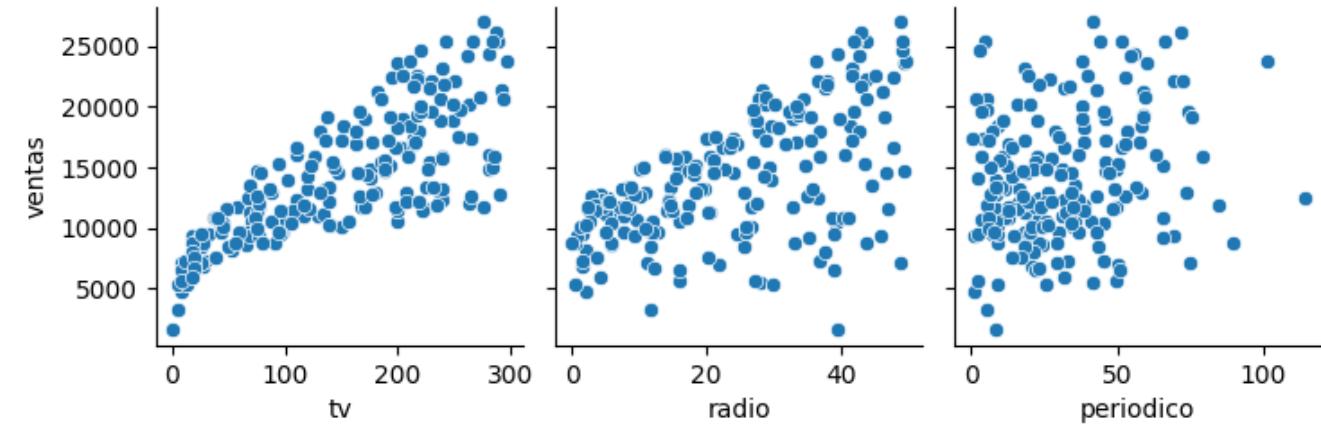
5. Predicción del Impacto de Estrategias Publicitarias:

Descripción: Evaluar cómo diferentes estrategias de marketing afectan las ventas y la percepción de marca.

Ejemplo: Determinar el aumento en ventas debido a una campaña publicitaria en redes sociales, considerando variables como el presupuesto y el alcance de la campaña.

Ejemplo

- Supongamos que somos consultores estadísticos contratados por un cliente para investigar la asociación entre el gasto en publicidad y las ventas de un producto en particular.
- Nuestro cliente no puede incrementar directamente las ventas del producto, pero puede controlar el gasto publicitario.
- Por lo tanto, si determinamos que existe una asociación entre publicidad y ventas, entonces podemos instruir a nuestro cliente para que ajuste los presupuestos publicitarios.



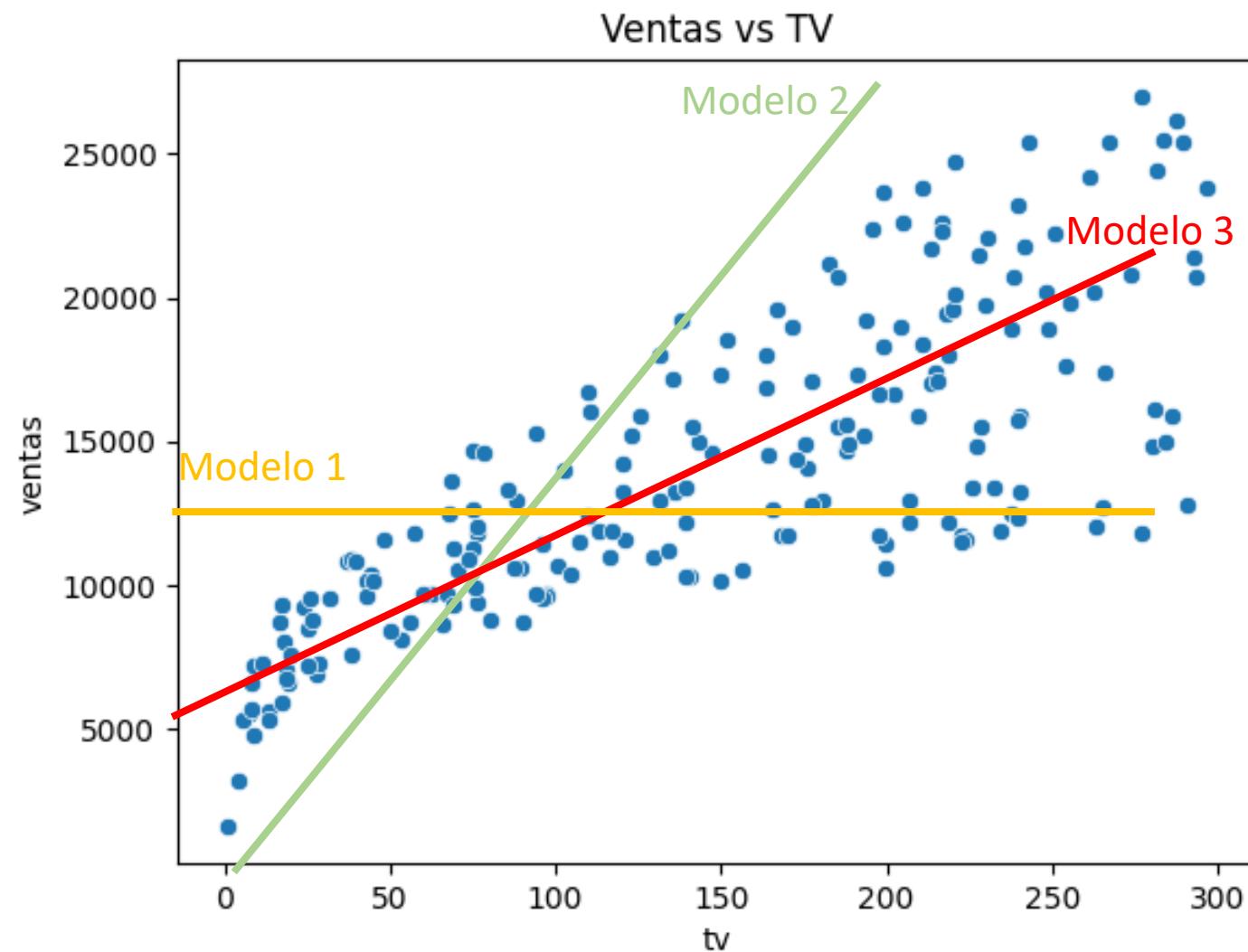
En otras palabras, nuestro objetivo es desarrollar un modelo preciso que se pueda utilizar para predecir las ventas sobre la base de los tres presupuestos de medios (TV, radio y periódico).

Con el modelo de regresión para predecir las ventas en función del presupuesto de publicidad en los diferentes medios, buscamos resolver las siguientes preguntas:

- ¿Existe una relación entre el presupuesto publicitario y las ventas?
- ¿Qué medios contribuyen a las ventas?
- ¿Qué contribución hace cada medio para el comportamiento de las ventas?
- ¿Con qué precisión podemos predecir las ventas futuras?
- ¿Existe correlación entre los medios publicitarios?

Regresión Lineal

- ¿Es posible trazar una línea que capture de forma perfecta todos los puntos?
- De forma intuitiva podemos decir que construir una modelo de predicción es encontrar la línea que nos permita capturar de la mejor manera la relación entre dos variables.
- En la gráfica podemos ver que el modelo 3, es el que mejor representa la relación entre las ventas y el presupuesto para TV.
- Observe que este modelo no captura de forma perfecta todos los puntos, por tanto la línea de tendencia central es el valor esperado de las ventas dado un presupuesto de TV.
- Debemos ser conscientes de que cualquier predicción de las ventas a partir del presupuesto para TV tendrá un error.
- No existe una predicción perfecta.



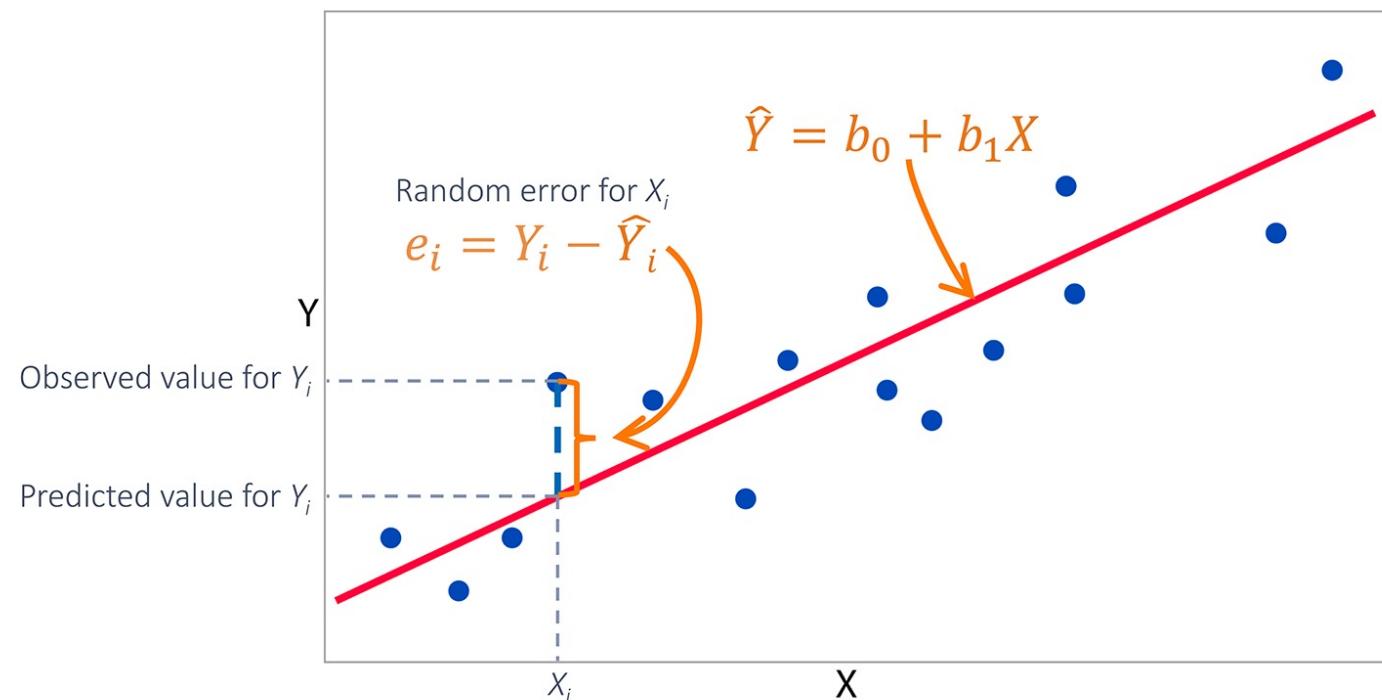
Conceptos Básicos

- **Objetivo o respuesta:** es la variable que deseamos predecir, en este caso las ventas. Generalmente nos referimos a la respuesta como la variable Y .
- **Entrada o predictor:** son las variables independientes o controlables que utilizamos para predecir la respuesta, nos referimos a estas como X 's, este caso los presupuestos de:
 - TV (X_1)
 - Radio (X_2)
 - Periódico (X_3)
- **Modelo de regresión:** es una función que relaciona la variable de respuesta con los predictores, generalmente se escribe como: $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$
- Donde, β_0 es el intercepto
- β_1, β_2 y β_3 son las contribuciones individuales (cuando las demás permanecen constantes) de cada variable de entrada a la predicción de la respuesta Y .
- **Error (ϵ):** es la diferencia entre el valor real de Y y su predicción \hat{Y} , por tanto
$$\rightarrow \epsilon = Y - \hat{Y}$$

Mínimos Cuadrados

- Construir un modelo de regresión implica encontrar las contribuciones de las variables TV, Radio y Periódico (β_1, β_2 y β_3) que permiten predecir las ventas con el mínimo error.
- Para estimar las constantes de la regresión se utiliza el método de mínimos cuadrados (Least Square).
- El método busca elegir los valores para $\beta_0, \beta_1, \beta_2$ y β_3 que minimicen la suma de cuadrados de los errores ϵ .
- En otras palabras:

$$\text{Min RSS} = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$



¿Qué contribución hace cada medio para la comportamiento de las ventas?

Al entrenar el modelo se obtiene la siguiente ecuación

$$Ventas = 2939 + 45.8TV + 188.5Radio - 1.0Periodico$$

- Las contribuciones de cada medio están expresadas en sus coeficientes:
 - TV:** por cada peso que se gasta en TV las ventas aumentan 45.8 unidades.
 - Radio:** por cada peso que se gasta en Radio las ventas aumentan 188.5 unidades.
 - Periódico:** por cada peso que se gasta en periódicos las ventas disminuyen en 1 unidad.
 - Intercepto:** cuando no se gasta en publicidad se espera que las ventas sean de 2938.9 unidades en promedio.

variables	coeficientes
Intercepto	2938.889369
tv	45.764645
radio	188.530017
periodico	-1.037493

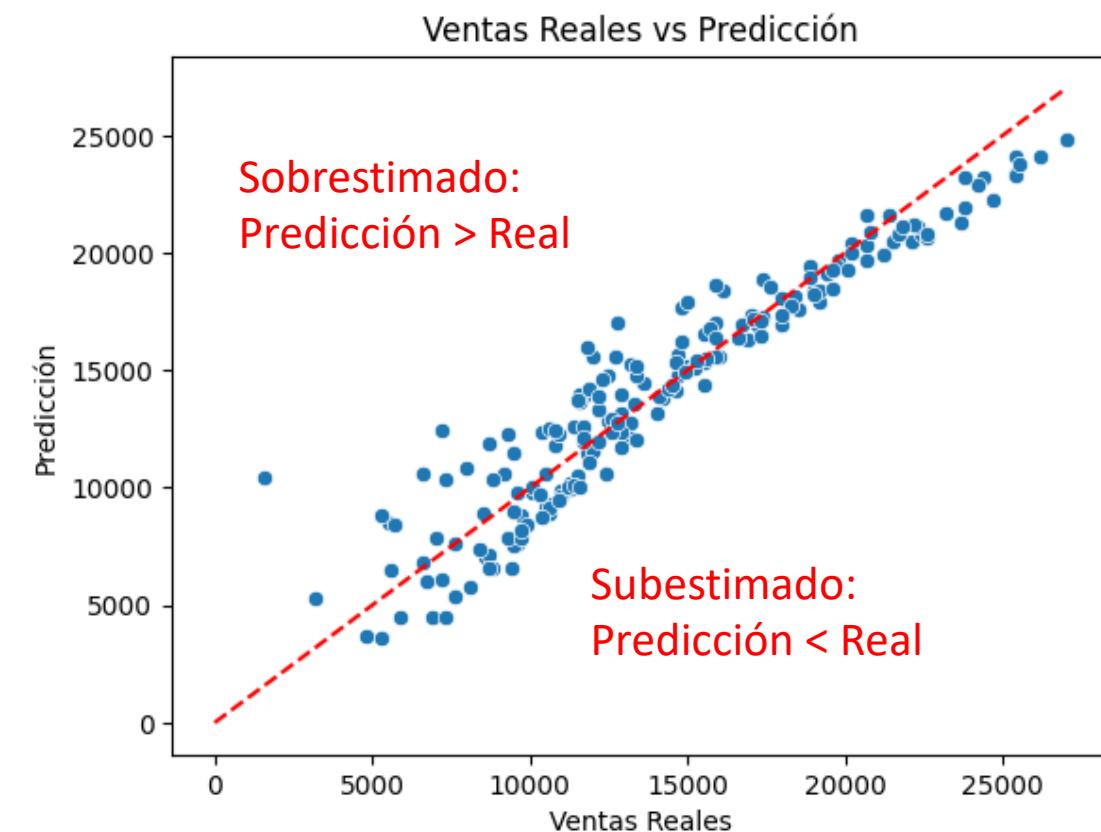
Análisis de las Predicciones

¿Con qué precisión podemos predecir las ventas futuras?

La tabla muestra la comparación entre las ventas reales y las predicciones

	Ventas Reales	Predictión	Error
0	22100.0	20524.0	1576.0
1	10400.0	12337.9	-1937.9
2	9300.0	12307.7	-3007.7
3	18500.0	17597.8	902.2
4	12900.0	13188.7	-288.7
5	7200.0	12478.3	-5278.3
6	11800.0	11729.8	70.2
7	13200.0	12123.0	1077.0

Las predicciones perfectas se darían sobre la línea roja



¿Cómo evaluar el modelo?

- Para evaluar las predicciones se utilizan indicadores que comparan los valores reales de las ventas con las predicciones que realizan los modelos o algoritmos.
- Como base se toma las diferencias entre el valor real de las ventas ($Real_i$) y el valor estimado para las ventas con el modelo lineal ($Predicción_i$)
- A continuación, se listan algunas métricas utilizadas:
 - Error medio absoluto (Mean Absolute Error):

$$MAE = \frac{\sum |Real_i - Predicción_i|}{n}$$

- Error porcentual absoluto medio MAPE (Mean Absolute Percentage Error):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Real_i - Predicción_i}{Real_i} \right|$$

Variables Categóricas

- Hasta el momento hemos asumido que todas las variables en nuestro modelo de regresión lineal son cuantitativas.
- Pero en la práctica algunos predictores son cualitativos.
- Si un predictor cualitativo (también conocido como factor), entonces incorporarlo a un modelo de regresión es muy simple, simplemente creamos un indicador o variable “dummy”

Factor					Variables Dummies			
TV	radio	periodico	ventas	zona	zona_norte	zona_occidente	zona_oriente	zona_sur
230.1	37.8	69.2	22.1	sur	0	0	0	1
44.5	39.3	45.1	10.4	occidente	0	1	0	0
17.2	45.9	69.3	9.3	occidente	0	1	0	0
151.5	41.3	58.5	18.5	occidente	0	1	0	0
180.8	10.8	58.4	12.9	oriente	0	0	1	0
...
38.2	3.7	13.8	7.6	occidente	0	1	0	0
94.2	4.9	8.1	9.7	sur	0	0	0	1
177.0	9.3	6.4	12.8	occidente	0	1	0	0
283.6	42.0	66.2	25.5	norte	1	0	0	0
232.1	8.6	8.7	13.4	norte	1	0	0	0



Pontificia Universidad
JAVERIANA
Bogotá

2.2 Arboles de Decisión

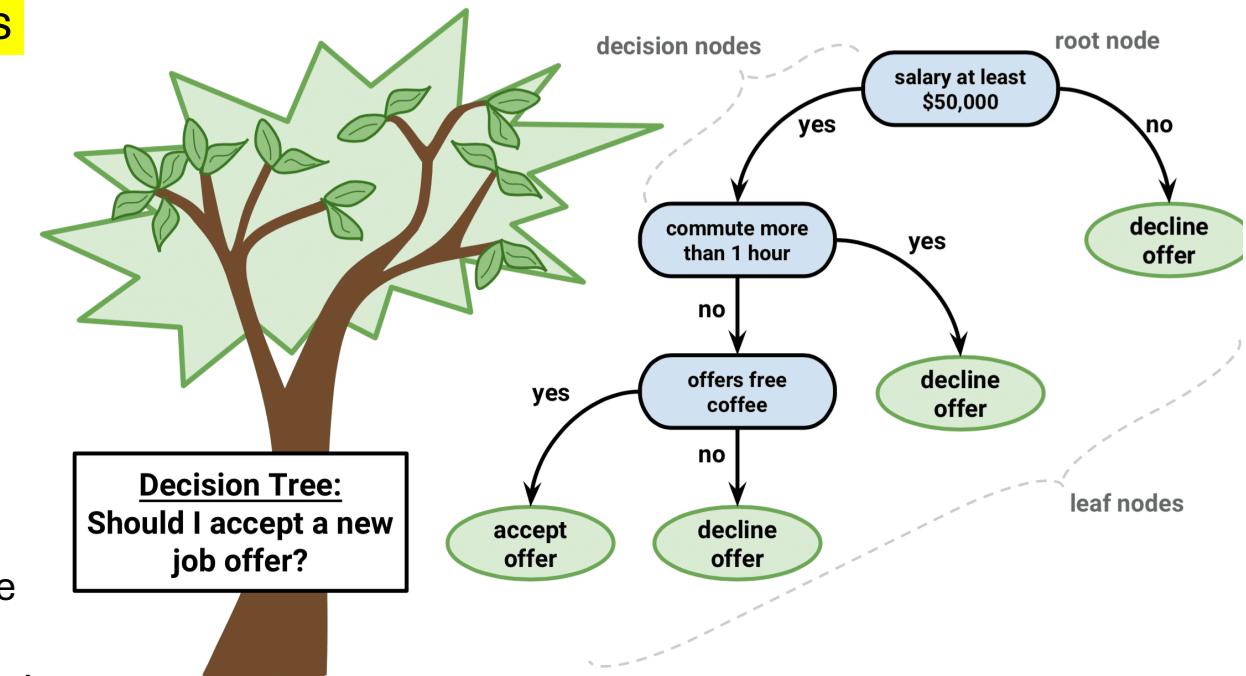
Educación **Continua**
Generamos experiencias educativas

Arboles de Decisión

Un árbol de decisión es un **conjunto de condiciones** organizadas en una **estructura jerárquica**, de tal manera que la **decisión final a tomar** se puede determinar **siguiendo las condiciones** que se cumplen **desde la raíz del árbol hasta sus hojas**.

De este árbol se pueden extraer las siguientes reglas (caminos)

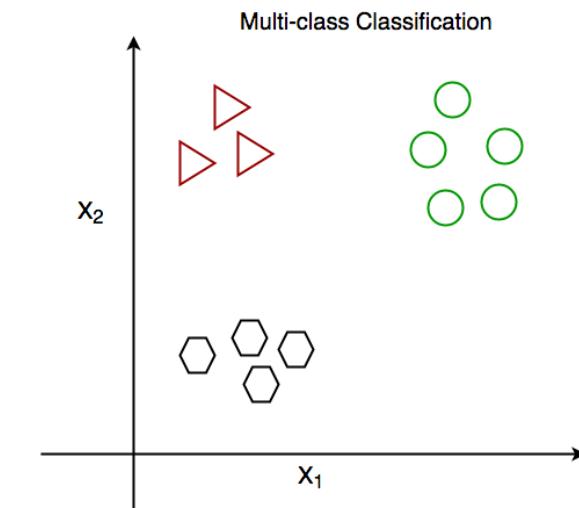
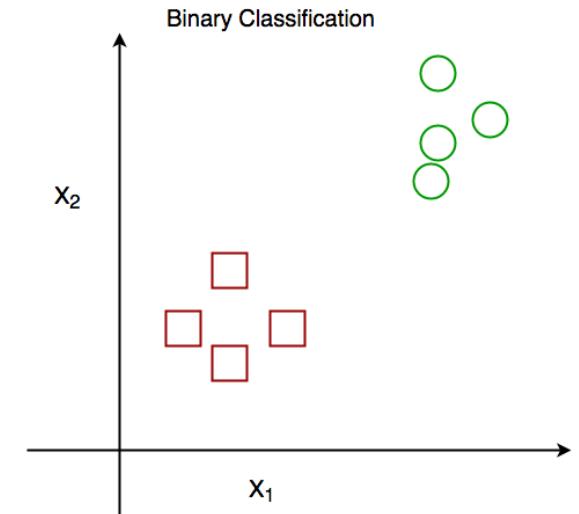
1. Salario < \$50.000 -> **Rechazar**.
2. Salario \geq \$50.000 + Transporte más de 1 hora -> **Rechazar**.
3. Salario \geq \$50.000 + Transporte menos de 1 hora + NO ofrece café gratis -> **Rechazar**.
4. Salario \geq \$50.000 + Transporte menos de 1 hora + ofrece café gratis -> **Aceptar**.



Usted puede construir un árbol de decisión a partir de sus propias reglas.
Sin embargo, esto puede que aplique solo para usted, y no es útil para generalizar en un grupo de interés.
En este curso nos enfocaremos en arboles de decisión que “aprende” reglas a partir de datos.

Problema de Clasificación

- El problema de clasificación consiste en predecir una clase o categoría
- Se dividen en:
 - *Problema Clasificación binaria* -> la variable a predecir tiene dos posibles resultados.
 - *Problema Multi-Clase* -> la variable a predecir tiene más de dos posibles resultados.
- Los modelos y algoritmos para predecir son llamados clasificadores.
- Algunos ejemplos de problemas de clasificación:
 - Determinar si un proyecto tendrá retrasos o sobrecostos
 - Predecir del impago (default) de un crédito
 - Determinar si un correo es spam
 - Determinar la condición médica de una persona de acuerdo a sus síntomas
 - Detectar transacciones fraudulentas
 - Clasificar imágenes de objetos



Ejemplo

Suponga que se desea construir un Árbol de Decisión para generar un sistema de recomendación de géneros de películas.

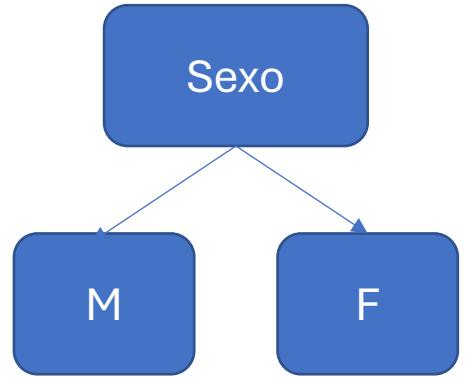
Se ha tomado una muestra de 10 personas a las que se registró:

1. Sexo
2. Edad
3. Genero de película de su preferencia

Sexo	Edad	Genero de Película
F	20	romántica
F	24	acción
M	25	acción
F	35	romántica
M	35	acción
M	40	romántica
F	40	romántica
F	50	documental
F	55	romántica
M	60	documental

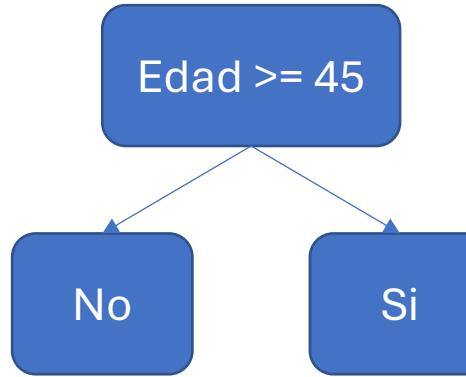
¿Cómo se Construye un Arbol de Decisión?

Divide y reinaras: Para todas las variables se generan divisiones de las clases y se calcula la **impureza** del nodo.



accion
accion
documental
romantica

accion
documental
romantica
romantica
romantica
romantica



accion
accion
accion
romantica

documental
documental
romantica

Al parecer la edad genera una división más “pura” (homogénea) de los géneros de películas.

Pero ¿Cómo medimos la pureza?

- Recuerde: $GINI_{nodo} = 1 - \sum_{k=1}^K p_k^2$
- Para la partición de la variable sexo:

$$GINI_{Sexo_M} = 1 - \left[\left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{2}{4}\right)^2 \right] = 0,625$$

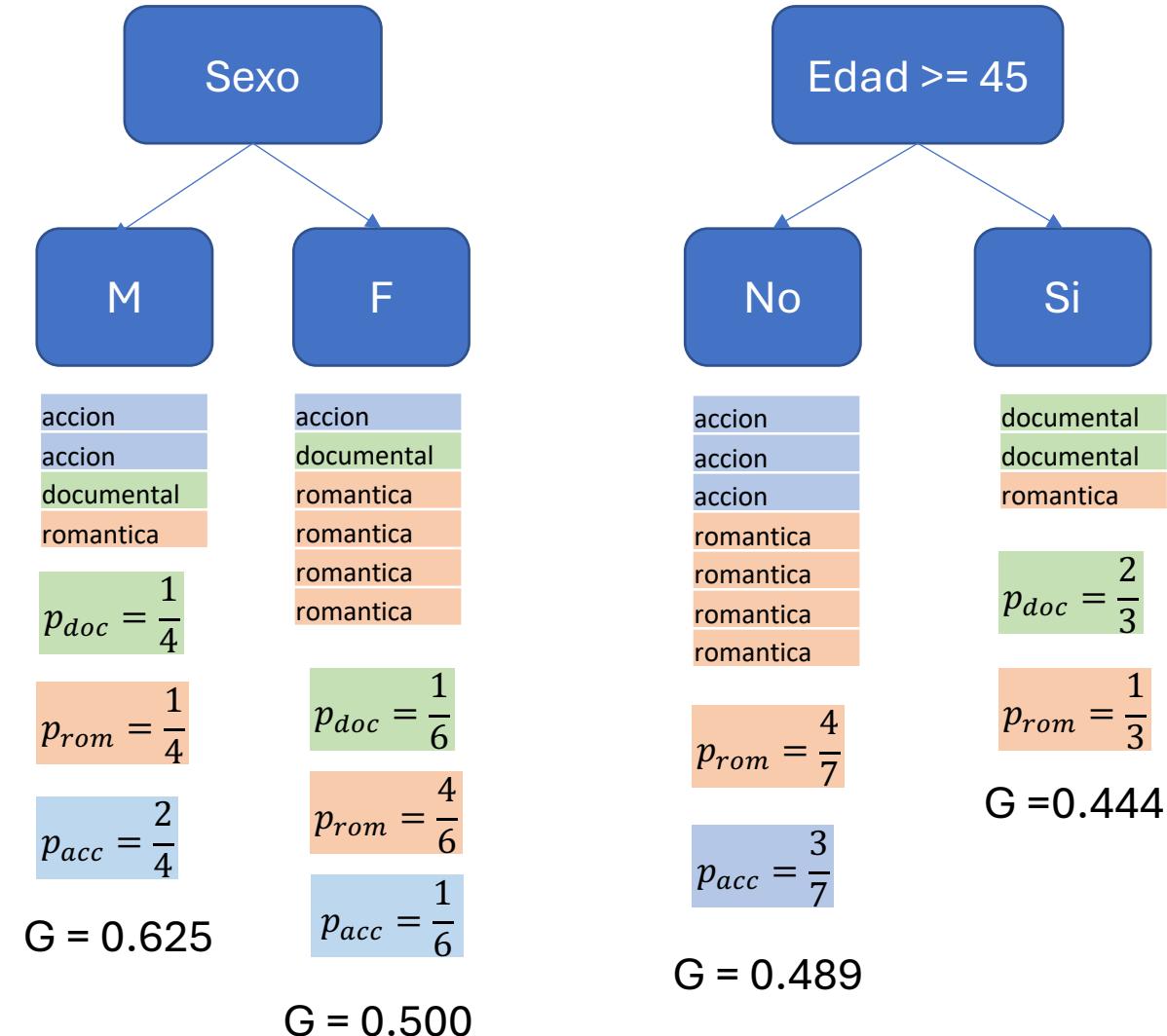
$$GINI_{Sexo_F} = 1 - \left[\left(\frac{1}{6}\right)^2 + \left(\frac{4}{6}\right)^2 + \left(\frac{1}{6}\right)^2 \right] = 0.500$$

- De esta manera:

$$GINI_{Sexo} = \left(\frac{4}{10}\right) 0,625 + \left(\frac{6}{10}\right) 0.500 = 0.550$$

$$GINI_{edad \geq 45} = \left(\frac{7}{10}\right) 0,489 + \left(\frac{3}{10}\right) 0.444 = 0.476$$

- Con este resultado la mejor partición (menor impureza) se genera con la partición edad ≥ 45 .
- A partir de los nodos generados se pueden generar más particiones teniendo en cuenta el mismo criterio.



Interpretación del árbol de decisión

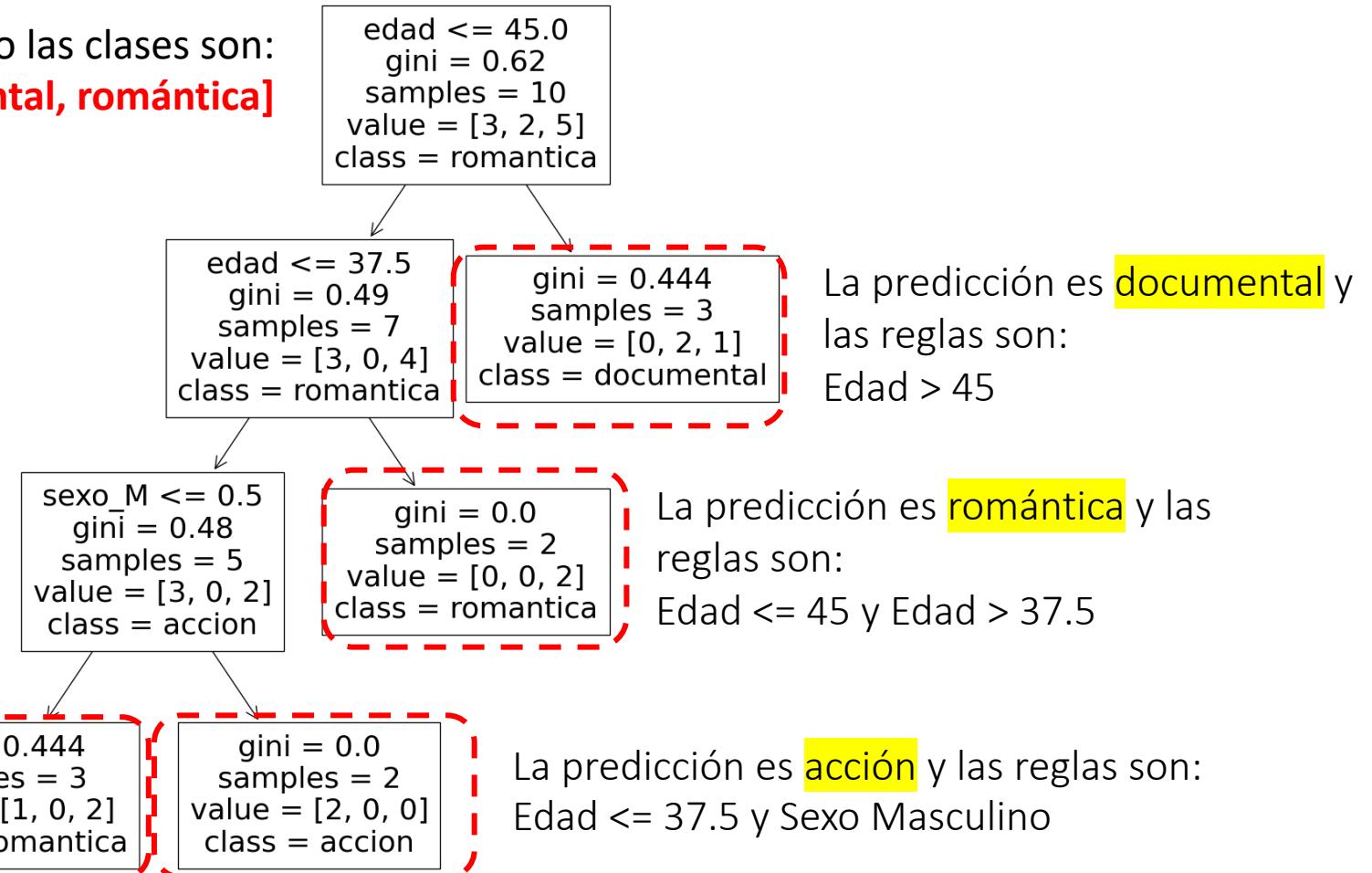
Suponga que se obtiene el siguiente árbol de decisión después calibrar el criterio de parada...

En cada nodo las clases son:
[acción, documental, romántica]

El árbol según la clase mayoritaria del nodo terminal.

En este nodo las clases son:
Acción: 1 voto (instancia)
Romántica: 2 votos (instancias)

La predicción es **romántica**
y las reglas son:
Edad <= 37.5 y Sexo Femenino



La predicción es **documental** y las reglas son:
Edad > 45

La predicción es **romántica** y las reglas son:
Edad <= 45 y Edad > 37.5

La predicción es **acción** y las reglas son:
Edad <= 37.5 y Sexo Masculino

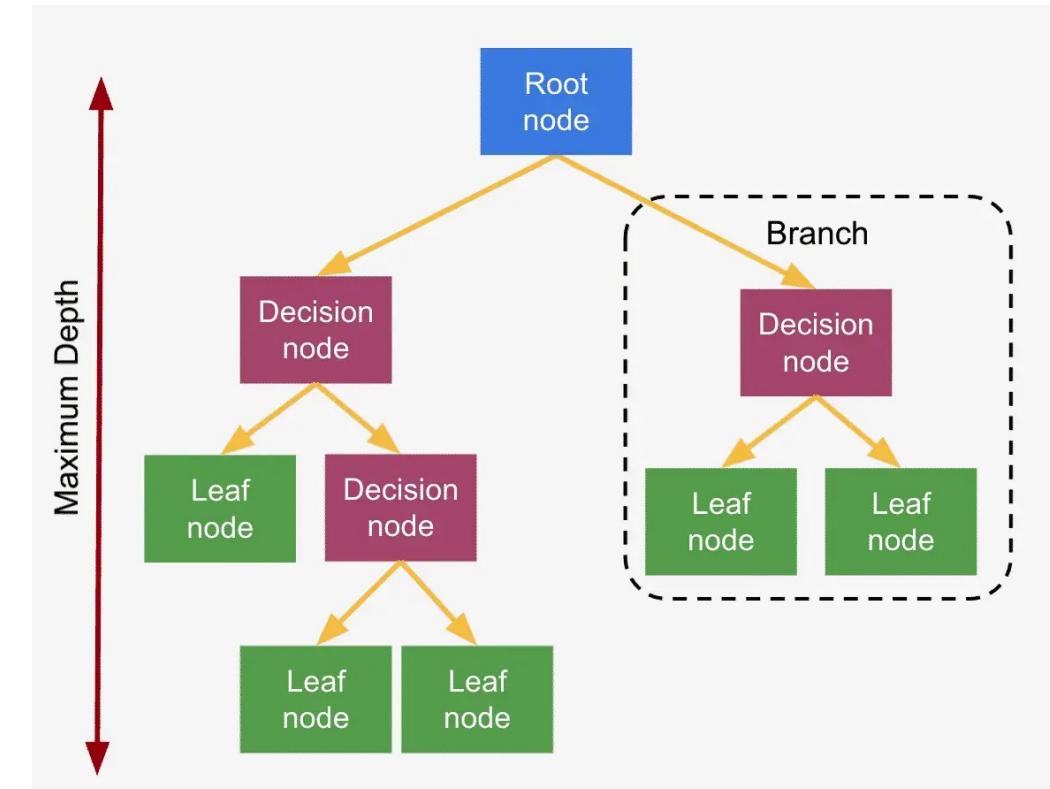
Criterios de parada

En resumen, un árbol de decisión se construye con los siguientes pasos:

1. Seleccionar una variable para dividir en el nodo actual (usando Gini, entropía u otros criterios).
2. Dividir el nodo en subnodos basados en los valores de la variable seleccionada.
3. Repetir el proceso en cada subnodo hasta que se cumpla un criterio de parada

Los criterios de parada para la construcción del árbol pueden ser:

- *Profundidad máxima (maximum depth)*
- Número mínimo de instancias por nodo (**min samples leaf**)
- El número mínimo de instancias necesarias para dividir un nodo interno (**min samples split**)
- Mínima disminución de impureza. Un nodo se dividirá si esta división induce una disminución de la impureza mayor o igual a este valor (**min impurity decrease**).



Matriz de Confusión

¿Qué tanto acierta el clasificador en las predicciones?

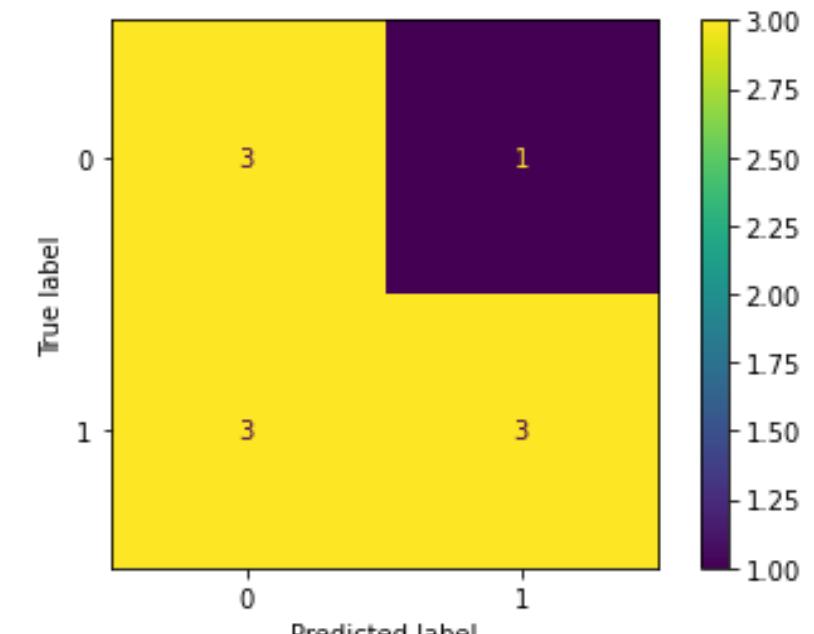
Para responder a esta pregunta se utiliza la [matriz de confusión](#), la cual resume los resultados de las predicciones.

		Predicción	
		1: Positivo	0: Negativo
Clase Real	1: Positivo	Verdadero Positivo VP	Falso Negativo FN
	0: Negativo	Falso Positivo FP	Verdadero Negativo VN

Medida de Desempeño	Formula
Exactitud (Accuracy): proporción de predicciones acertadas	Exactitud = $\frac{VP + VN}{VP + FP + FN + VN}$
Recuperación (Recall): proporción predicciones positivas correctas, con respecto al total de clases reales positivas.	Recuperación = $\frac{VP}{VP + FN}$
Precisión : proporción de predicciones positivas correctas, con respecto al total de predicciones.	Precisión = $\frac{VP}{VP + FP}$
F-1 Score: se calcula como la media harmónica entre el Recall y la Precisión.	F1-Score = $2 \frac{\text{Precisión} \times \text{Recuperación}}{\text{Precisión} + \text{Recuperación}}$

Matriz de confusión y métricas

Cliente	Gasto Anual	Tarjeta	Utiliza Cupón	Probabilidad	Predictión
12	1.92	0	0	0.0502	0
11	5.55	0	0	0.3895	0
4	3.92	0	1	0.1725	0
9	1.18	1	1	0.4961	0
27	1.27	1	0	0.5116	1
25	2.64	1	1	0.7284	1
20	6.28	0	1	0.5129	1
29	4.80	0	0	0.2760	0
10	3.35	0	1	0.1235	0
31	5.53	1	1	0.9512	1

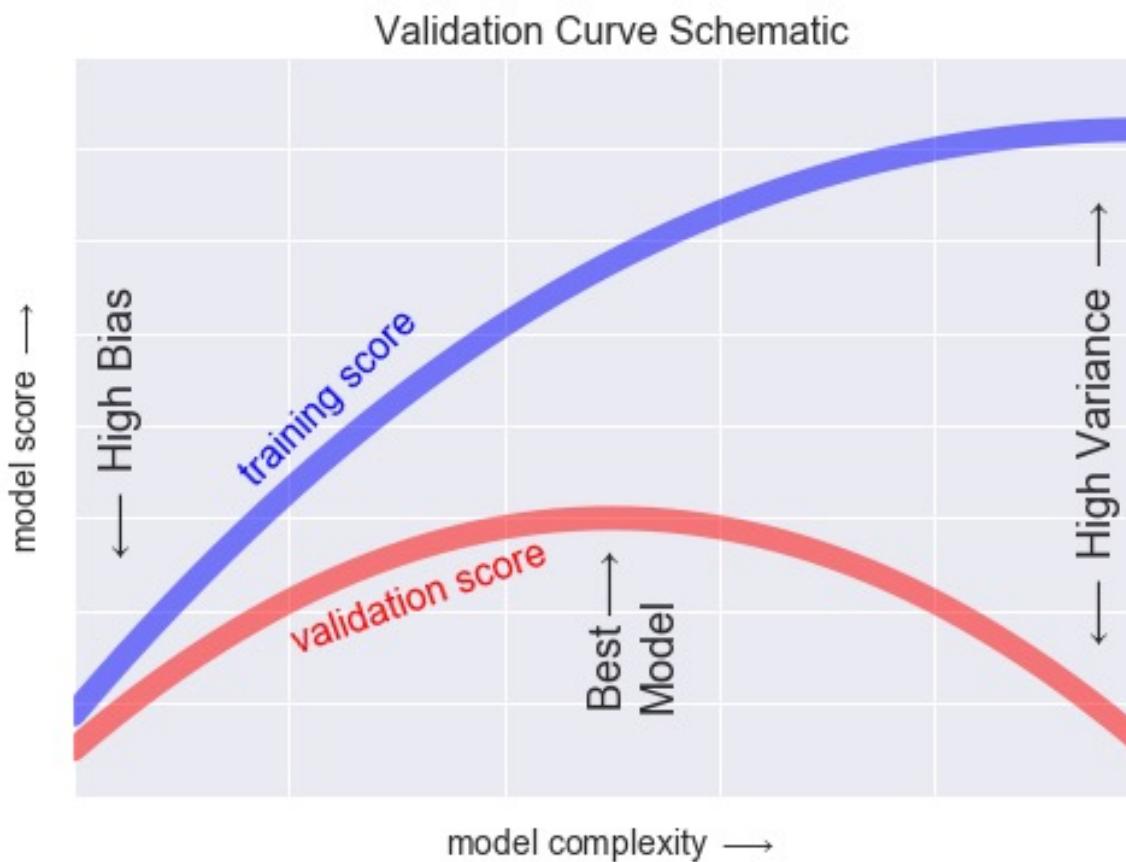


$$Accuracy = \frac{3 + 3}{3 + 1 + 3 + 3} = 0.6 \text{ En solo en 6 de las 10 predicciones el modelo acertó (60%)}$$

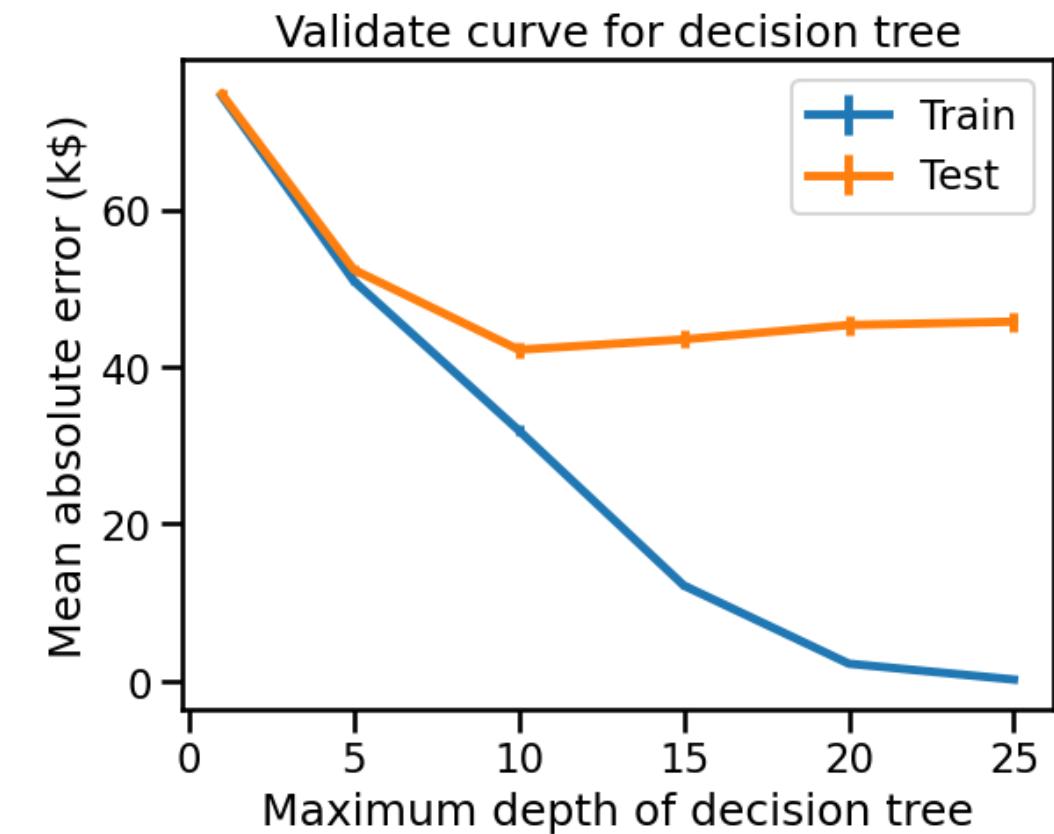
$$Sensibilidad = \frac{3}{3 + 3} = 0.5 \text{ De los 6 clientes que usaron el cupón, el modelo detectó 3 (50%)}$$

$$Precisión = \frac{3}{3 + 1} = 0.75 \text{ El modelo predijo que 4 clientes usarían el cupón, pero solo 3 lo hicieron (75%)}$$

Curvas de validación



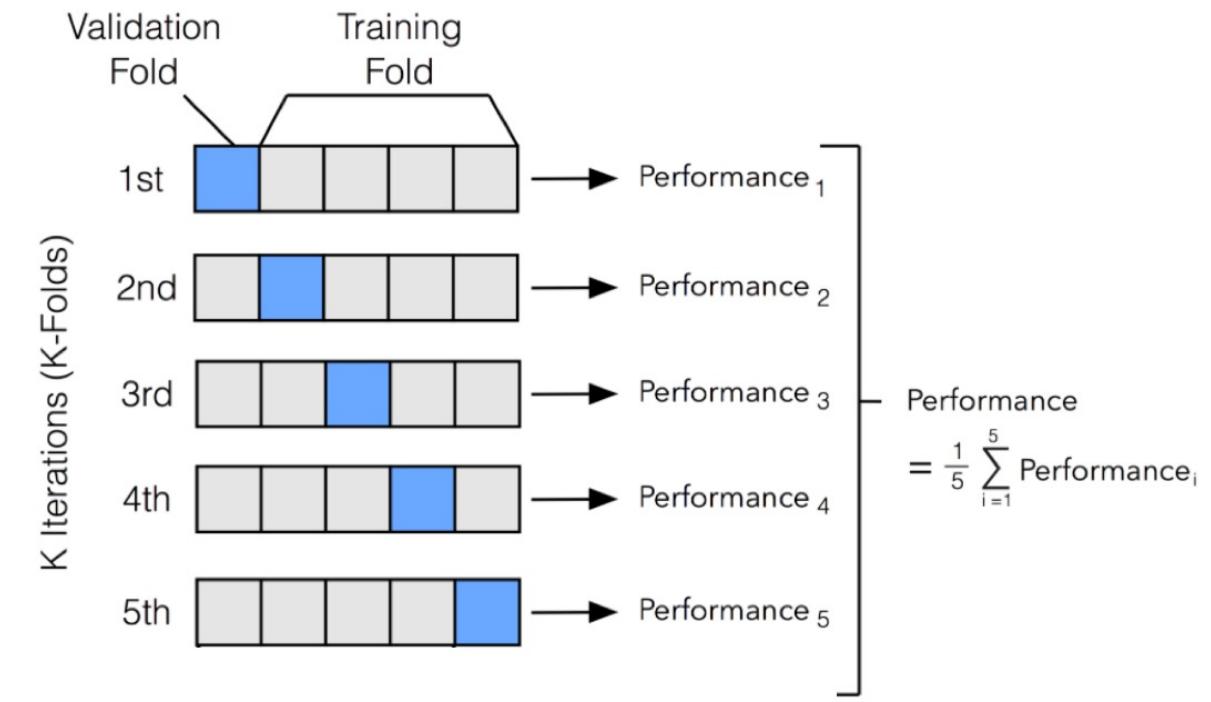
El hiperparámetro óptimo se obtiene cuando comienza a haber divergencia en la curva de entrenamiento y validación.



En este ejemplo observe que al aumentar la profundidad del árbol después de 10, el MSE no disminuye. Por tanto, una profundidad de 10 puede ser un valor óptimo del hiperparámetro.

Validación Cruzada

- En general se utiliza una técnica de remuestreo para estimar el rendimiento de los modelos y ajustar hiperparámetros.
- Se utiliza un subconjunto de muestras para ajustar un modelo y las muestras restantes se utilizan para estimar la eficacia del modelo.
- Este proceso se repite varias veces y los resultados se agregan y resumen.
- En la validación cruzada las muestras se dividen aleatoriamente en k conjuntos de igual tamaño.
- Se ajusta un modelo utilizando todas las muestras excepto un subconjunto, el cual se utiliza para validación.





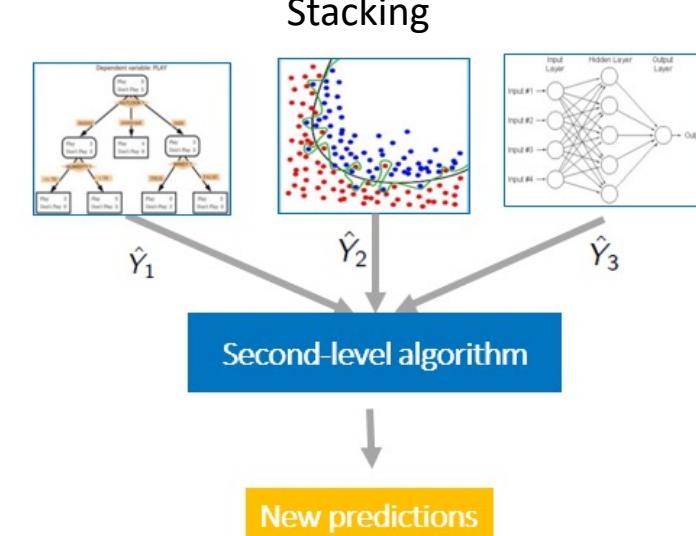
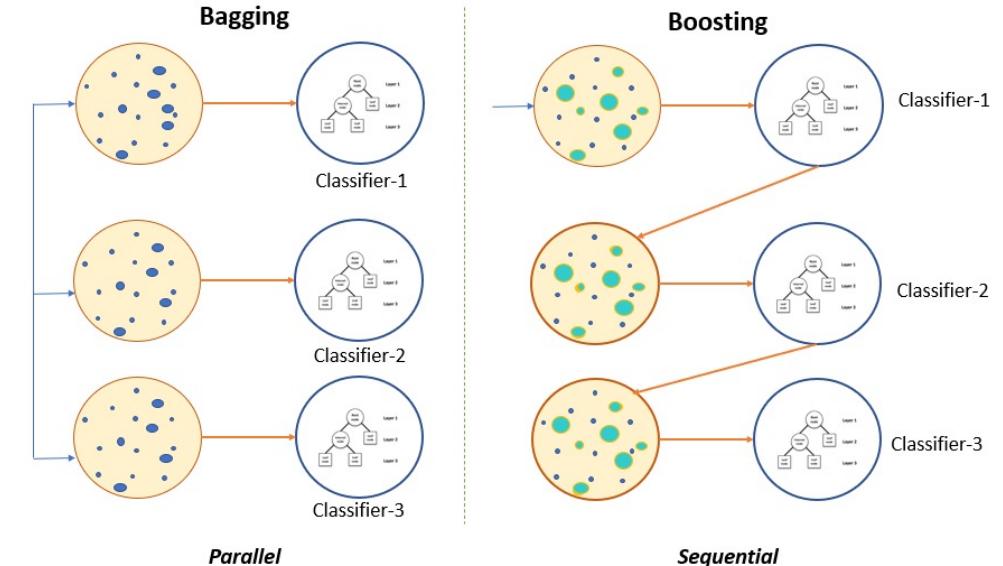
Pontificia Universidad
JAVERIANA
Bogotá

2.3 Métodos de Ensamble

Educación **Continua**
Generamos experiencias educativas

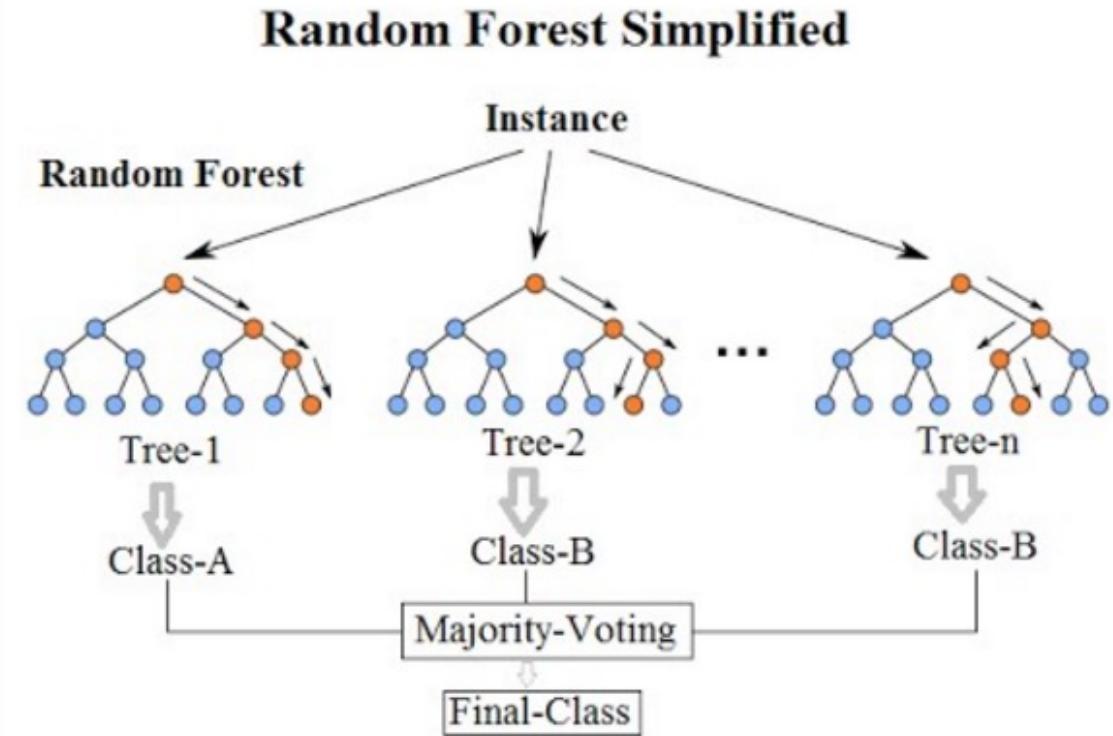
Métodos de Ensamble

- Un método de ensamble es un enfoque que combina muchos modelos simples.
- La idea es tener "bloques de construcción" para obtener un modelo único y potencialmente muy poderoso.
- Estos modelos de bloques de construcción simples a veces se conocen como aprendices débiles, ya que pueden conducir a predicciones mediocres por sí mismos.
- Estos son métodos de conjunto para los cuales el bloque de construcción puede ser una simple regresión o un árbol de clasificación.
- Los enfoques de ensamble mas utilizados son:
 - Bagging (Bootstrap Aggregation)
 - Boosting
 - Stacking



Bagging – Random Forest

- Los árboles de decisión sufren de una alta variabilidad.
- Esto significa que, si dividimos los datos de entrenamiento en dos partes y ajustamos un árbol de decisión a ambas mitades, los resultados que obtengamos podrían ser bastante diferentes.
- Por el contrario, un procedimiento con poca varianza arrojará resultados similares si se aplica repetidamente a distintos conjuntos de datos.
- El Bagging es un procedimiento de propósito general para reducir la varianza de un método de aprendizaje estadístico.
- El Random Forest es el algoritmo más conocido que funciona bajo el método Bagging.
- Con el Random Forest se generan una cantidad determinada arboles a partir de una selección aleatoria de variables e instancias.
- EL número de árboles, la cantidad de variables e instancias son parámetros de calibración.



La predicción final se da por promedio para la regresión o por votación mayoritaria para un problema de clasificación.