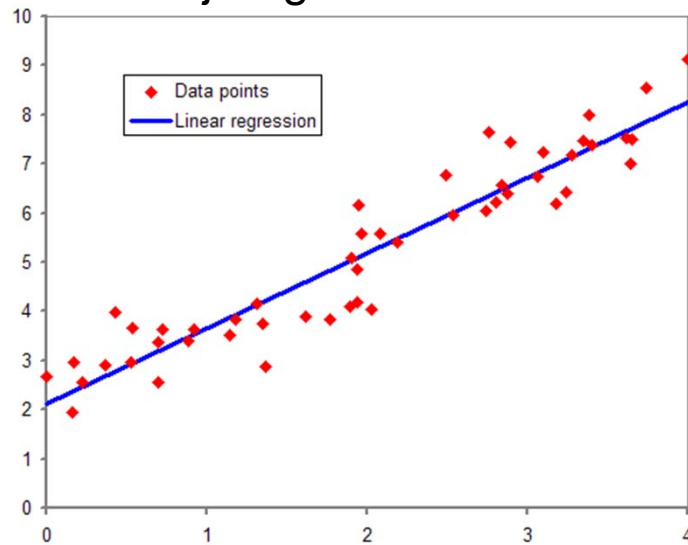


REGRESIÓN

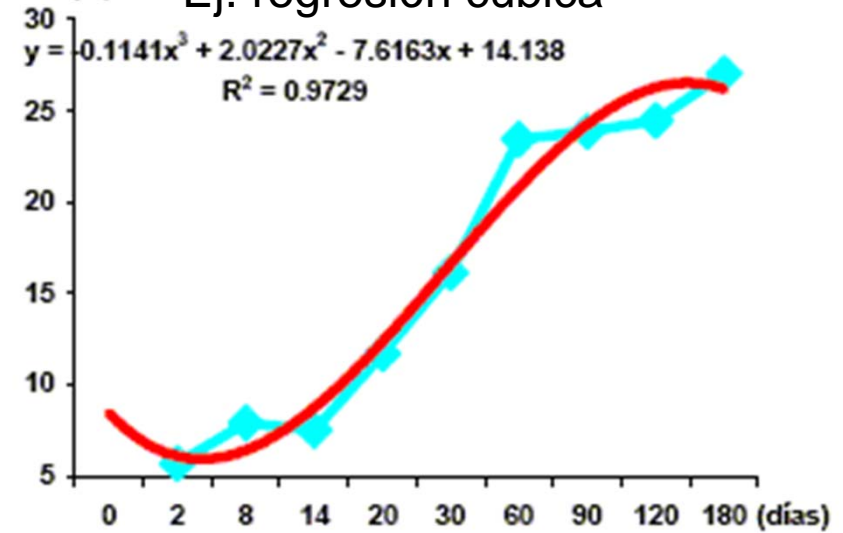
- Método que modela la relación entre una variable dependiente Y y unas independientes X_i
 - Regresión lineal: intentar aproximar una función $f(x)$ (supuestamente lineal) con una función lineal $f'(x)$
$$f'(x) = w_0 + w_1 a_1(x) + w_2 a_2(x) + \dots + w_n a_n(x)$$
 - No tiene por qué ser lineal:
 - Cuadrática
 - Cúbica
 - Cuártica
 - Logarítmica
 - Logística
 - etc.

REGRESIÓN

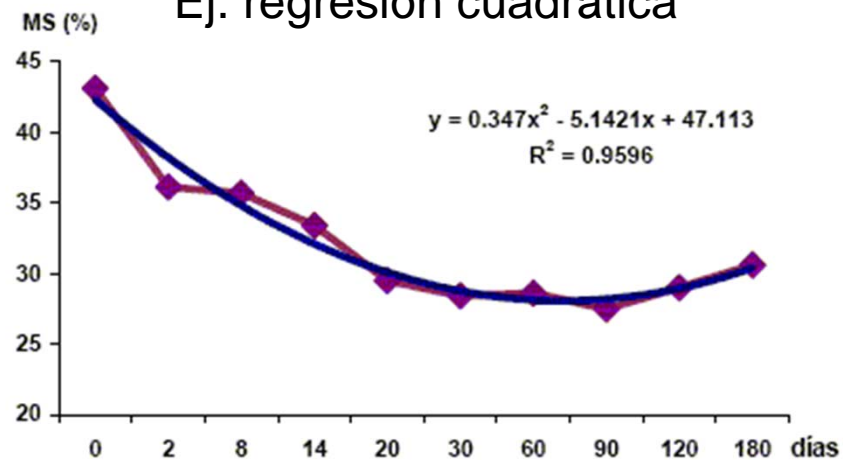
Ej. regresión lineal



Ej. regresión cúbica



Ej. regresión cuadrática



REGRESIÓN

- Objetivo de la regresión:
 - Minimizar el error entre la función aproximada y el valor de la aproximación
- Una posible medida de error es la suma del error cuadrático medio sobre el conjunto de entrenamiento total, D:

$$E = \frac{1}{2} \sum_{x \in D} (f(x) - f'(x))^2$$

REGRESIÓN

- En regresión lineal, el problema de definir la función

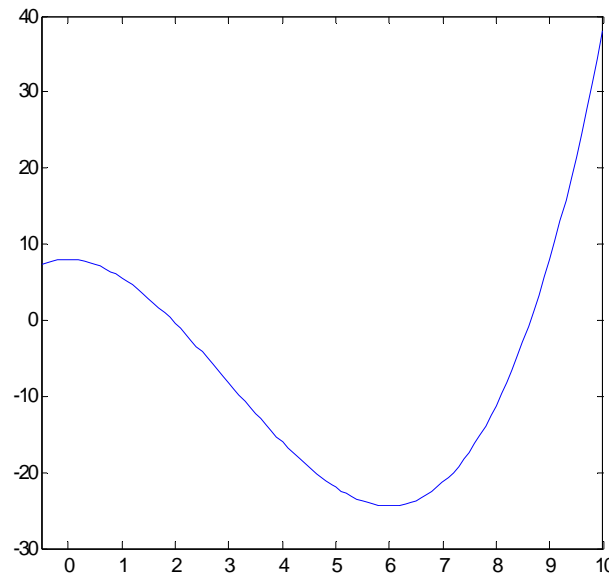
$$f'(x) = w_0 + w_1 a_1(x) + w_2 a_2(x) + \dots + w_n a_n(x)$$

se traslada a un problema de definir el vector de pesos w

- Se debe encontrar el vector w que minimice la función de error:
 - Problema de búsqueda en el espacio de pesos
- Métodos estadísticos conocidos
 - Mínimos cuadrados

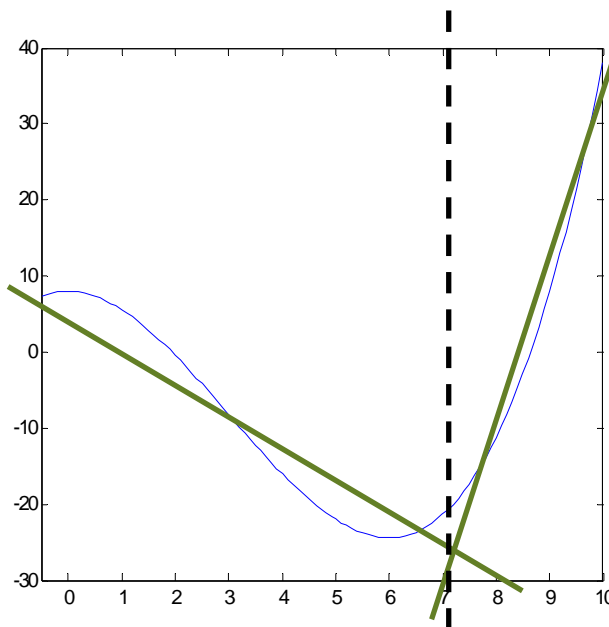
REGRESIÓN

- Árboles de regresión y de modelos de regresión:
 - La regresión lineal es un modelo global
 - Hay una ecuación aplicable (**modelo global**) en todo el espacio
 - Si datos interactúan de forma no lineal y complicada, conseguir un modelo lineal global puede ser arduo
 - Ejemplo: ¿Modelo lineal?



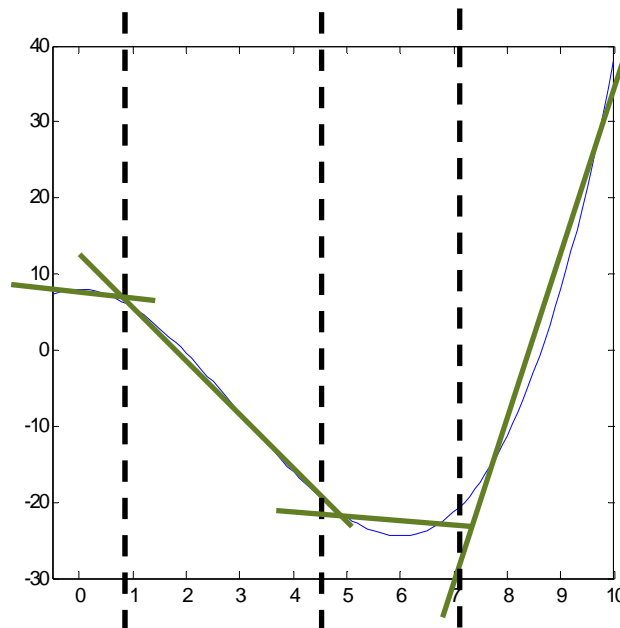
REGRESIÓN

- Árboles de regresión y de modelos de regresión:
 - Alternativa: partir el espacio en regiones más pequeñas
 - En estas regiones, las interacciones serán más manejables
 - En cada región, crear un modelo sencillo
 - **Modelo local**
 - Ejemplo:



REGRESIÓN

- Árboles de regresión y de modelos de regresión:
 - Alternativa: partir el espacio en regiones más pequeñas
 - Partir las regiones recursivamente
 - Ejemplo:



REGRESIÓN

- Árboles de regresión y de modelos de regresión:
 - ¿Cómo partir el espacio?
 - Árboles: similar a los árboles de clasificación
 - ¿Qué modelo crear en las regiones?
 - Árboles de regresión
 - Árboles de modelos de regresión

ÁRBOLES DE REGRESIÓN

- Árboles de regresión:
 - Generalización de los árboles de decisión para dar funciones de regresión continuas
 - Hojas:
 - Valor medio de la respuesta de las instancias que llegan a esa hoja
 - Sistema CART (*Classification And Regression Trees*) (Breiman et al., 1984)
 - Tanto para árboles de clasificación como de regresión
 - Sirve para todo tipo de atributos
 - Son un caso particular de los árboles de modelos de regresión

ÁRBOLES DE REGRESIÓN

- Árboles de regresión:
 - Aproximan la función objetivo mediante una función **constante a trozos**:
 - Es decir, el modelo en cada región es una constante
 - Si a un nodo hoja llega un conjunto de c instancias (x_i, y_i) , el modelo en esa hoja será la el valor promedio de los valores de la variable dependiente de las instancias que lleguen a ese nodo:

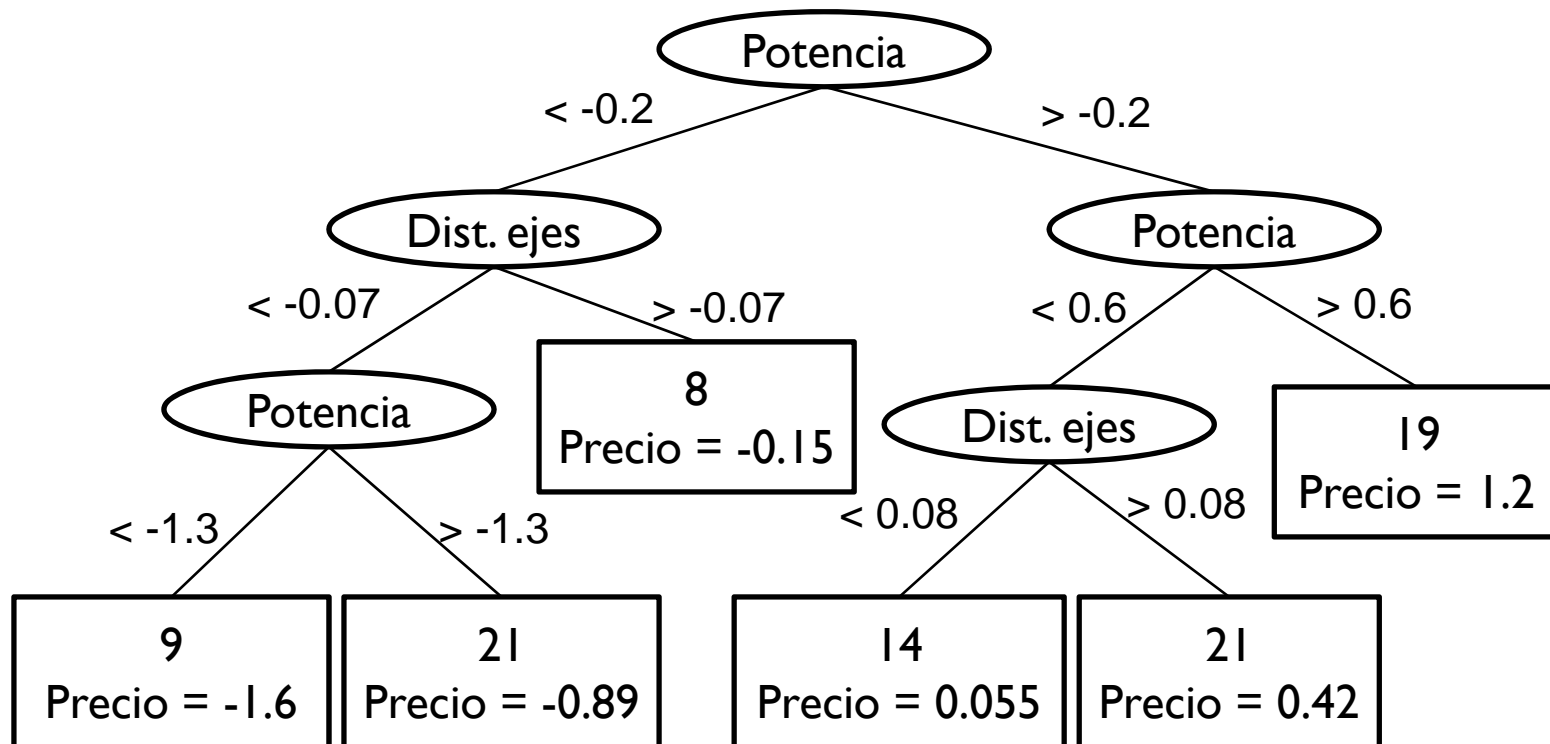
$$\frac{1}{c} \sum_{i=1}^c y_i$$

ÁRBOLES DE REGRESIÓN

- Árboles de regresión:
 - CART:
 - Criterio de impureza:
 - En lugar de utilizar el índice de Gini, el criterio de impureza es la **suma de cuadrados**
 - Por tanto, se escogen las divisiones que causan la mayor reducción en la suma de los cuadrados
 - Poda:
 - Al podar el árbol la medida utilizada es el error cuadrático medio en las predicciones hechas por el árbol

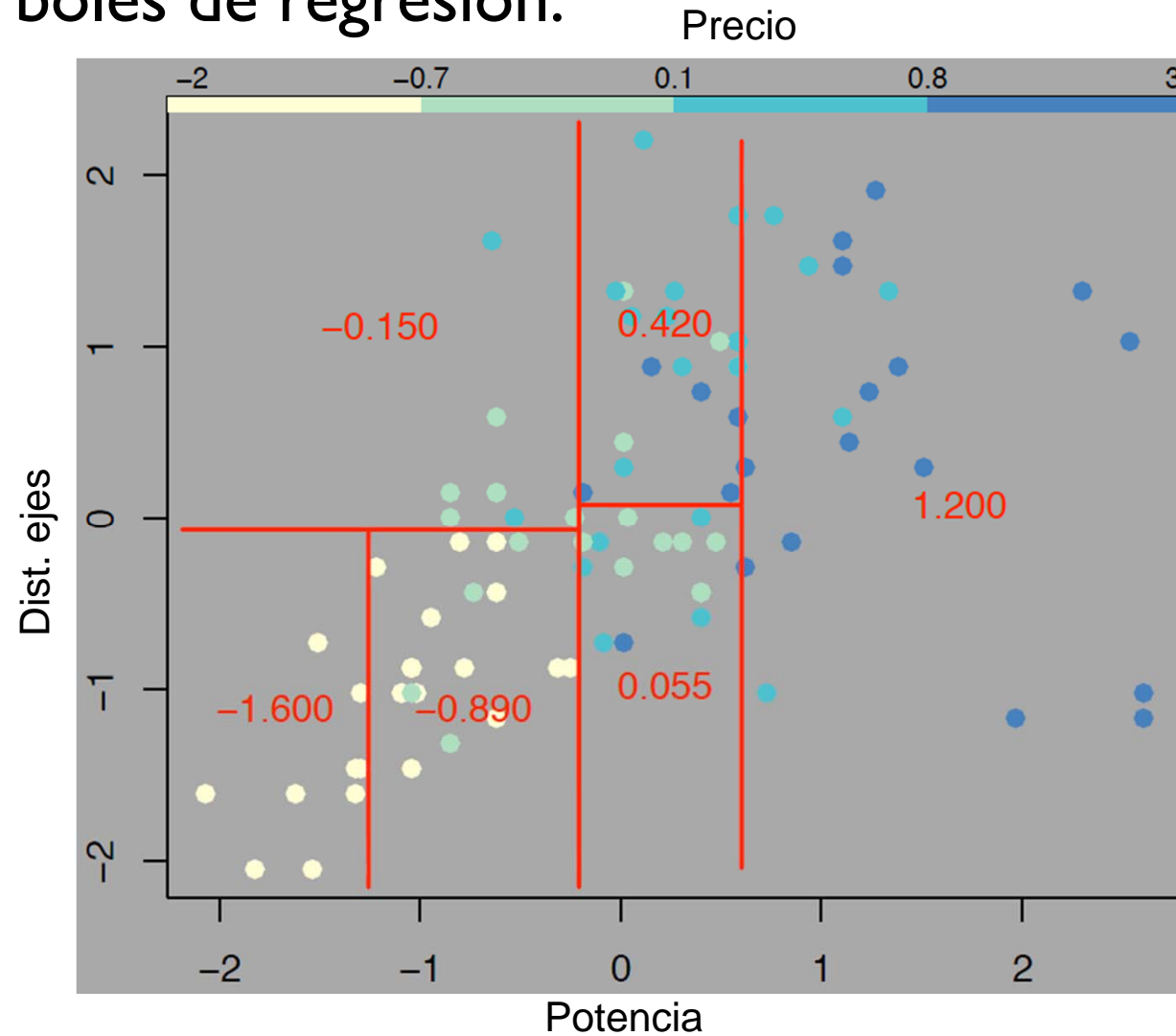
ÁRBOLES DE REGRESIÓN

- Árboles de regresión:
 - Ejemplo: Precios de coches de 1993:
 - Todos los atributos, y la variable de salida, han sido normalizados
 - Media cero y desviación típica 1



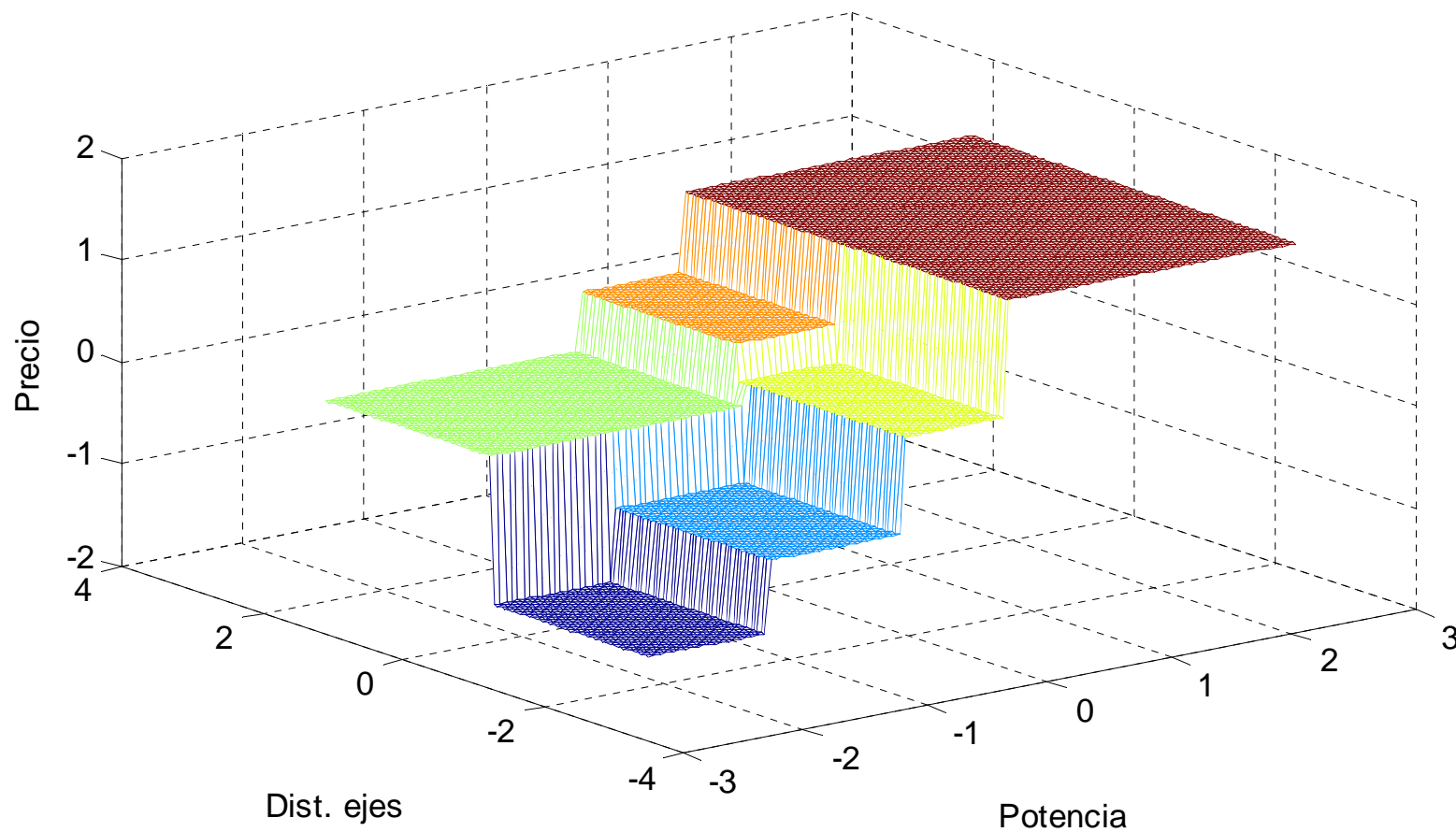
ÁRBOLES DE REGRESIÓN

- Árboles de regresión:



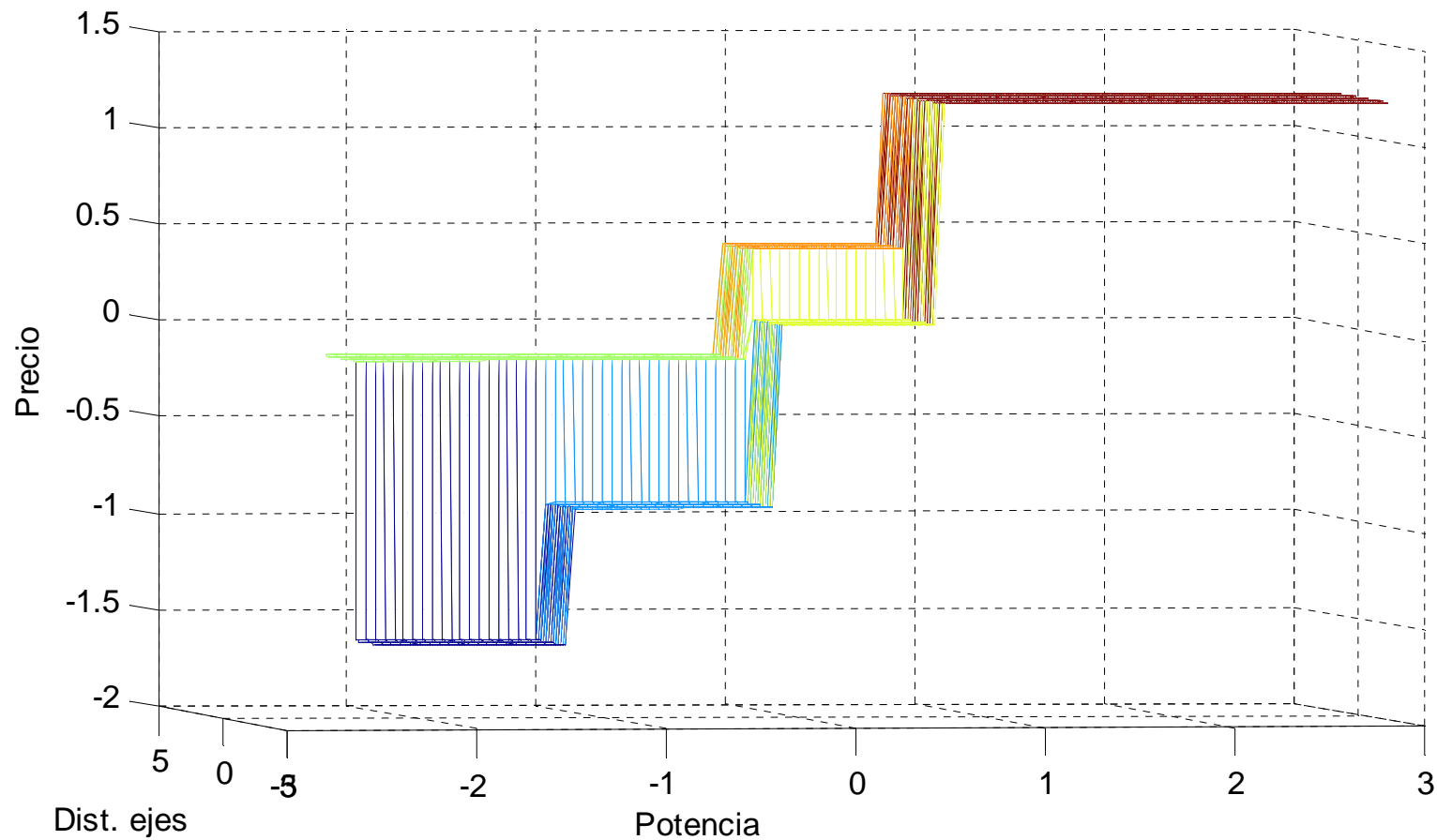
ÁRBOLES DE REGRESIÓN

- Árboles de regresión:



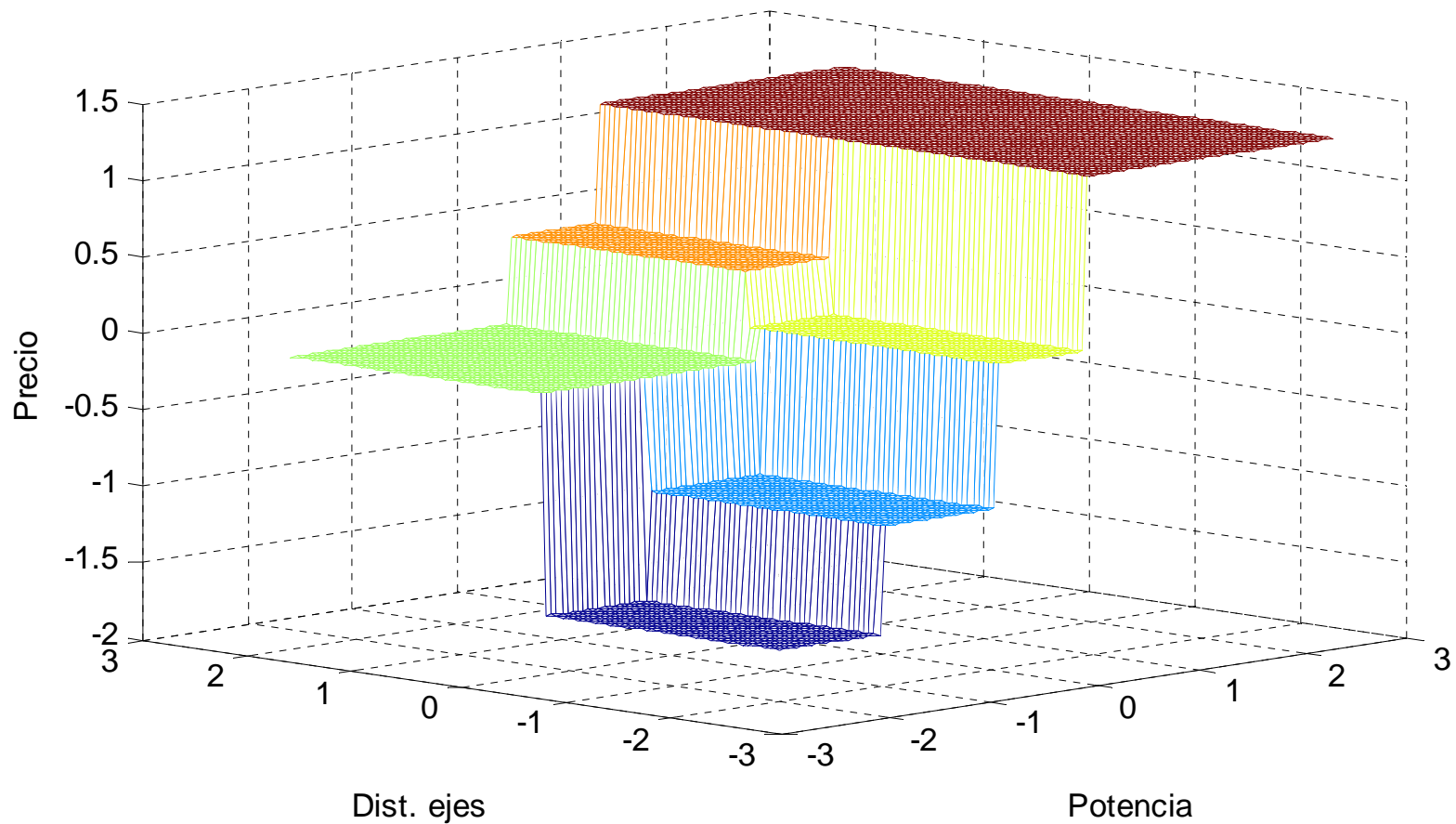
ÁRBOLES DE REGRESIÓN

- Árboles de regresión:



ÁRBOLES DE REGRESIÓN

- Árboles de regresión:



ÁRBOLES DE REGRESIÓN

- Árboles de regresión:

- Base de datos con 209 configuraciones diferentes de ordenadores

	Cycle time (ns) MYCT	Main memory (KB)		Cache (KB) CACH	Channels		Performance PRP
		Min. MMIN	Max. MMAX		Min. CHMIN	Max. CHMAX	
1	125	256	6000	256	16	128	198
2	29	8000	32000	32	8	32	269
3	29	8000	32000	32	8	32	220
4	29	8000	32000	32	8	32	172
5	29	8000	16000	32	8	16	132
...							
207	125	2000	8000	0	2	14	52
208	480	512	8000	32	0	0	67
209	480	1000	4000	0	0	0	45

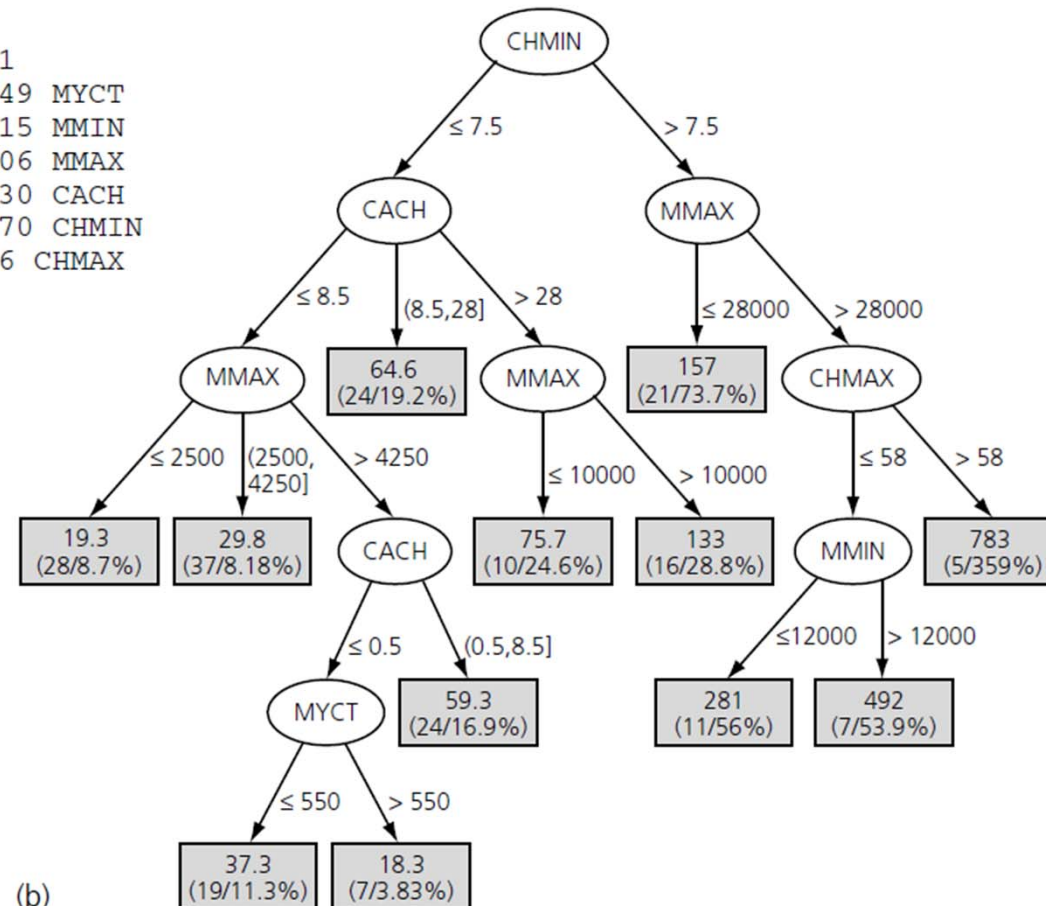
- Salida: potencia de procesamiento
- Todos los atributos, y la salida, son continuos

ÁRBOLES DE REGRESIÓN

- Árboles de regresión:
 - Ecuación de regresión (lineal) (a) y árbol de regresión (b)

PRP =
-56.1
+0.049 MYCT
+0.015 MMIN
+0.006 MMAX
+0.630 CACH
-0.270 CHMIN
+1.46 CHMAX

(a)



(b)

ÁRBOLES DE REGRESIÓN

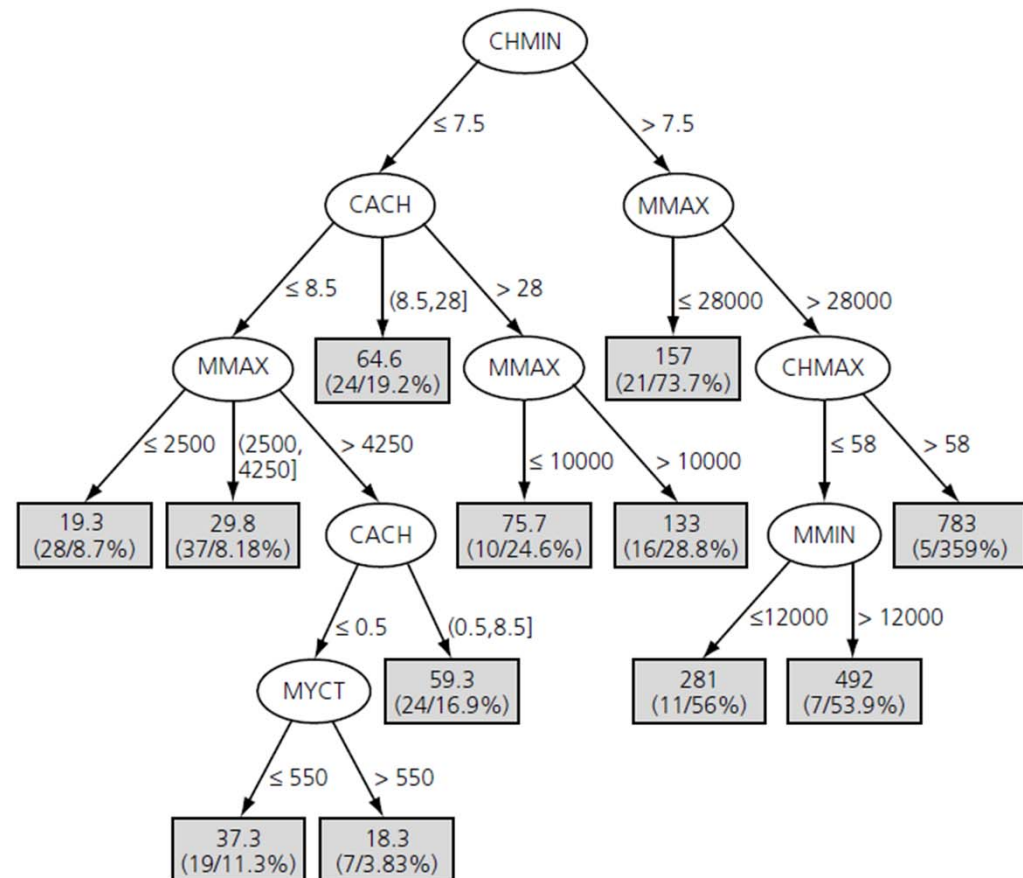
- Árboles de regresión:
 - Ventajas:
 - Realizar predicciones es rápido
 - Es fácil comprender qué variables son importantes a la hora de realizar la predicción
 - Si falta algún dato,
 - Puede no ser posible bajar por todo el árbol hasta la hoja
 - Aún así, se puede hacer una predicción promediando todas las hojas del subárbol al que se llega
 - Por ejemplo:

ÁRBOLES DE REGRESIÓN

- Árboles de regresión:

- Ejemplo:

- CHMIN = 8
 - CHMAX = ?
 - MMIN = 8000
 - MMAX = 32000



ÁRBOLES DE REGRESIÓN

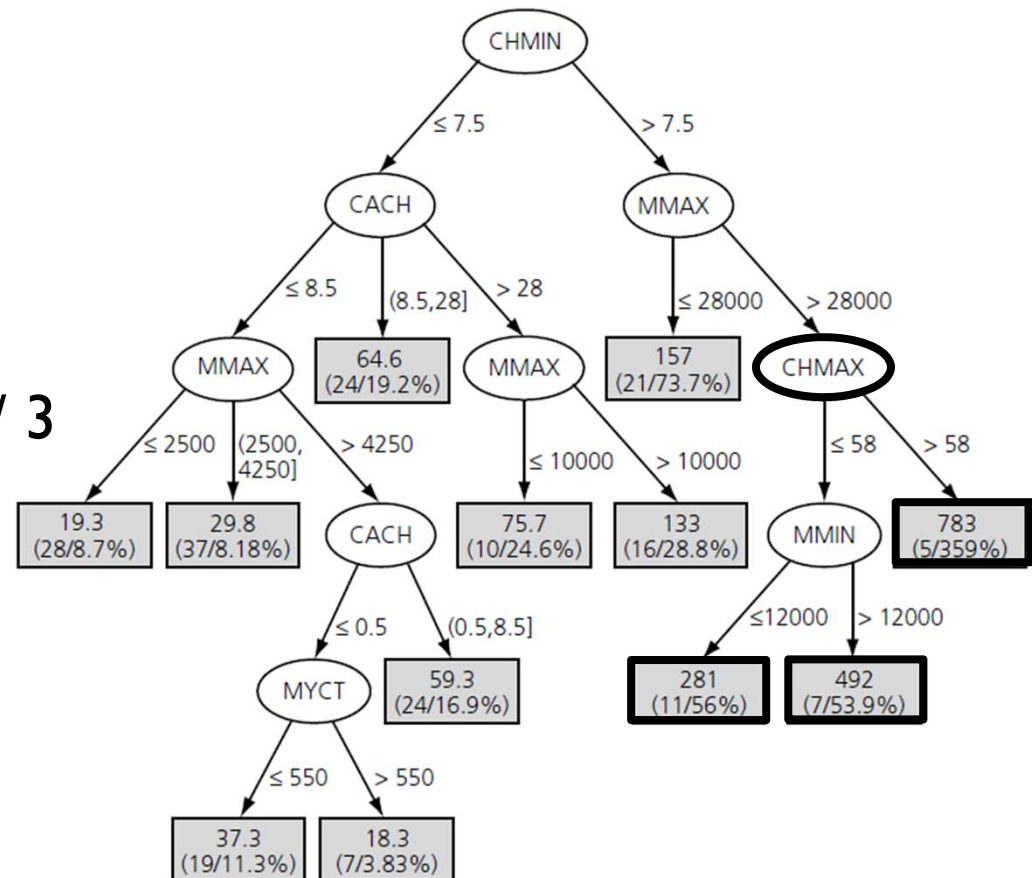
- Árboles de regresión:

- Ejemplo:

- CHMIN = 8
- CHMAX = ?
- MMIN = 8000
- MMAX = 32000

- Predicción:

- $(281 + 492 + 783) / 3$



ÁRBOLES DE REGRESIÓN

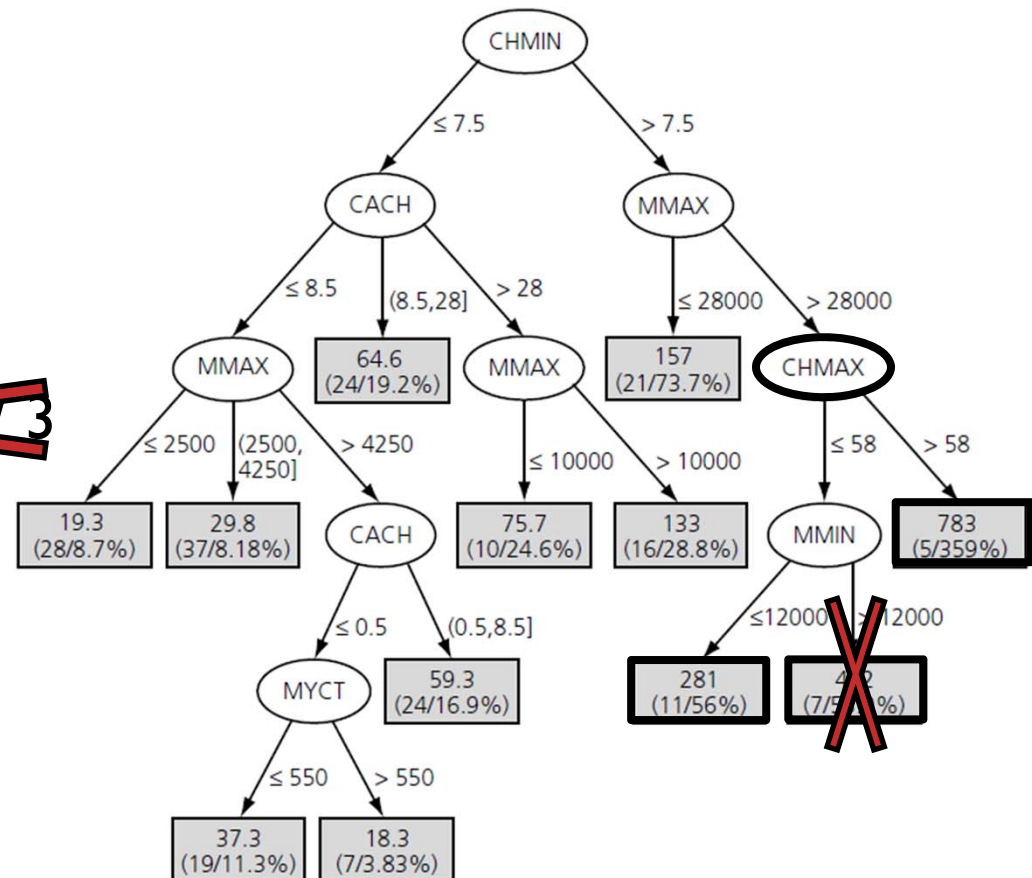
- Árboles de regresión:

- Ejemplo:

- CHMIN = 8
- CHMAX = ?
- MMIN = 8000
- MMAX = 32000

- Predicción:

- ~~(281 + 422 + 783) / 3~~



ÁRBOLES DE REGRESIÓN

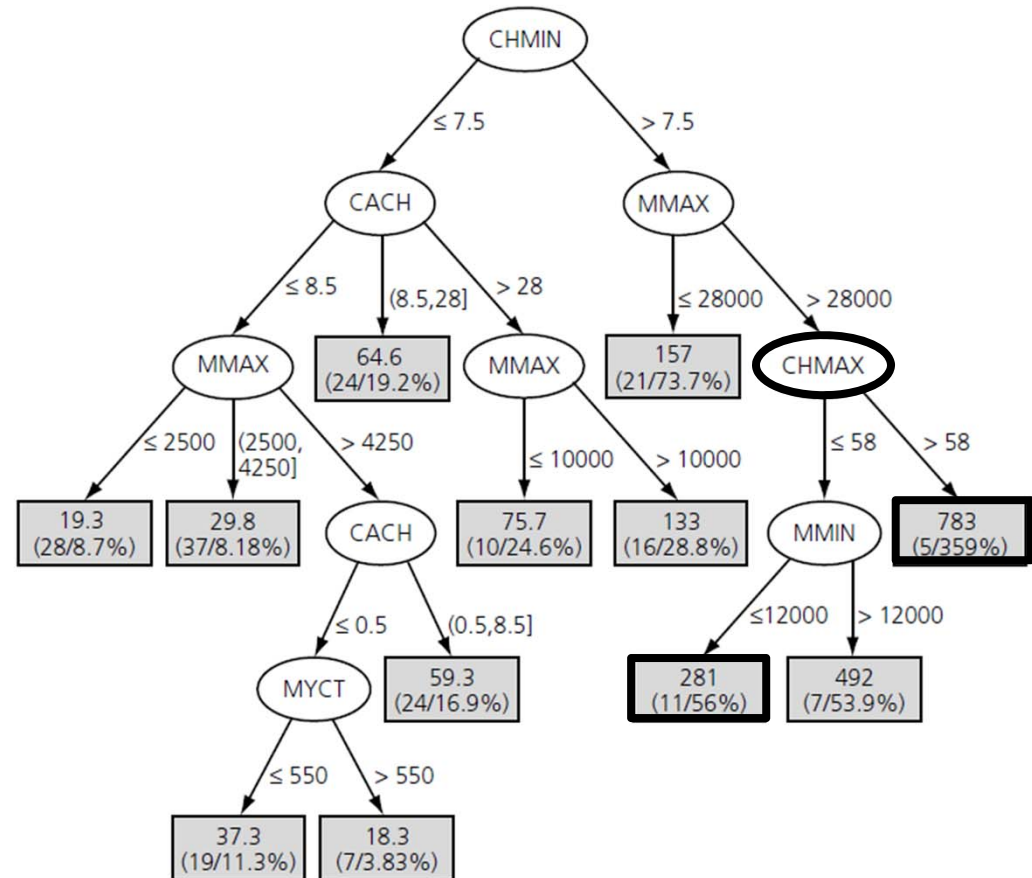
- Árboles de regresión:

- Ejemplo:

- CHMIN = 8
 - CHMAX = ?
 - MMIN = 8000
 - MMAX = 32000

- Predicción:

- $(281 + 783) / 2$



ÁRBOLES DE REGRESIÓN

- Árboles de regresión:
 - Ventajas:
 - Realizar predicciones es rápido
 - Es fácil comprender qué variables son importantes a la hora de realizar la predicción
 - Si falta algún dato,
 - Puede no ser posible bajar por todo el árbol hasta la hoja
 - Aún así, se puede hacer una predicción promediando todas las hojas del subárbol al que se llega
 - El modelo da una respuesta escalonada
 - Se puede aplicar cuando la superficie de regresión deseada es suave
 - Aún así, si no es suave, la superficie constante a trozos puede aproximarla arbitrariamente bien
 - Con un número suficiente de hojas
 - Hay algoritmos rápidos y eficientes para desarrollar estos árboles

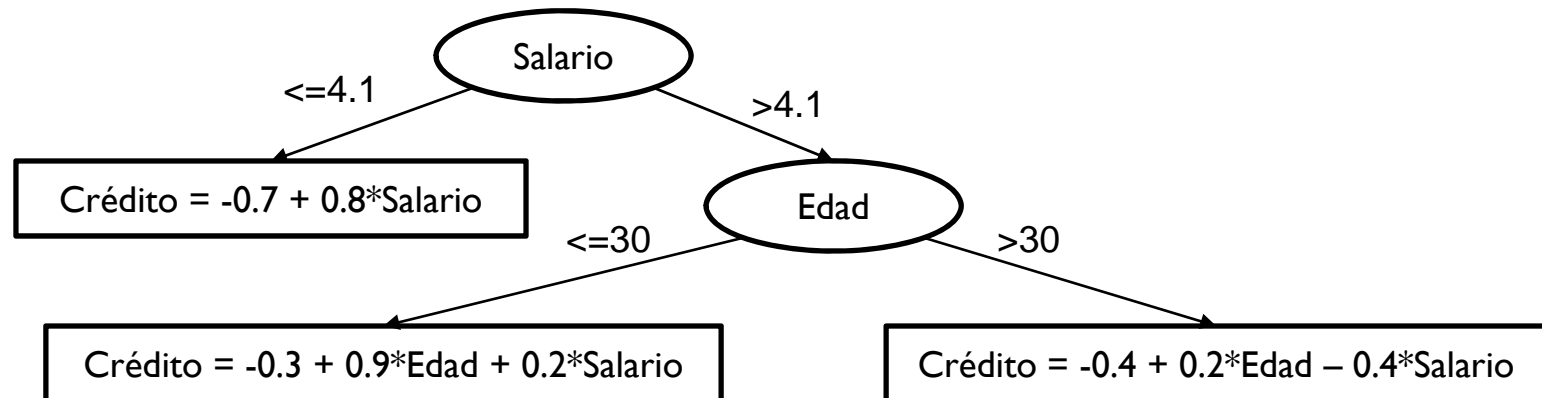
ÁRBOLES DE REGRESIÓN

- Árboles de modelos de regresión:
 - Hojas:
 - Contienen un modelo de regresión lineal para predecir la respuesta de las instancias que llegan a esa hoja
 - Algoritmo M5 (Quinlan, 1992) y M5' (Wang y Witten, 1997)
 - M5 genera árboles de decisión similares a los de ID3
 - Escoge el atributo que minimiza la variación en los valores de la clase que resulta por cada rama
 - M5' introduce mejoras para:
 - Tratar atributos discretos
 - Tratar valores perdidos
 - Reducir el tamaño del árbol

ÁRBOLES DE REGRESIÓN

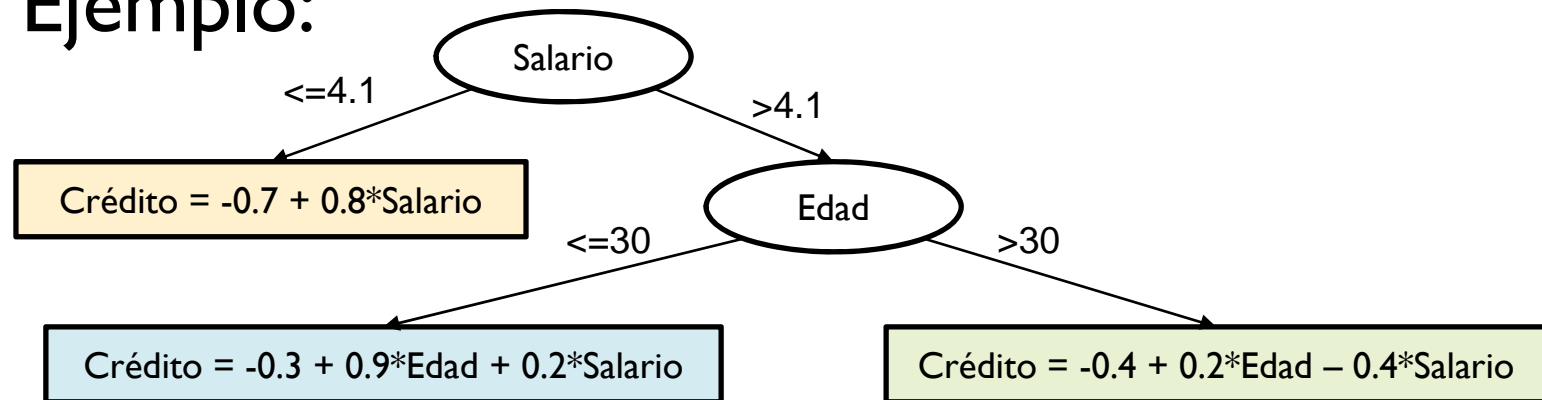
- Ejemplo:

Cliente	Salario	Edad	Hijos	...	Crédito
1	4.5	34	1	...	4.53
2	2.3	27	2	...	1.16
3	9.5	51	2	...	6.21
4	1.2	29	3	...	0.13
...	



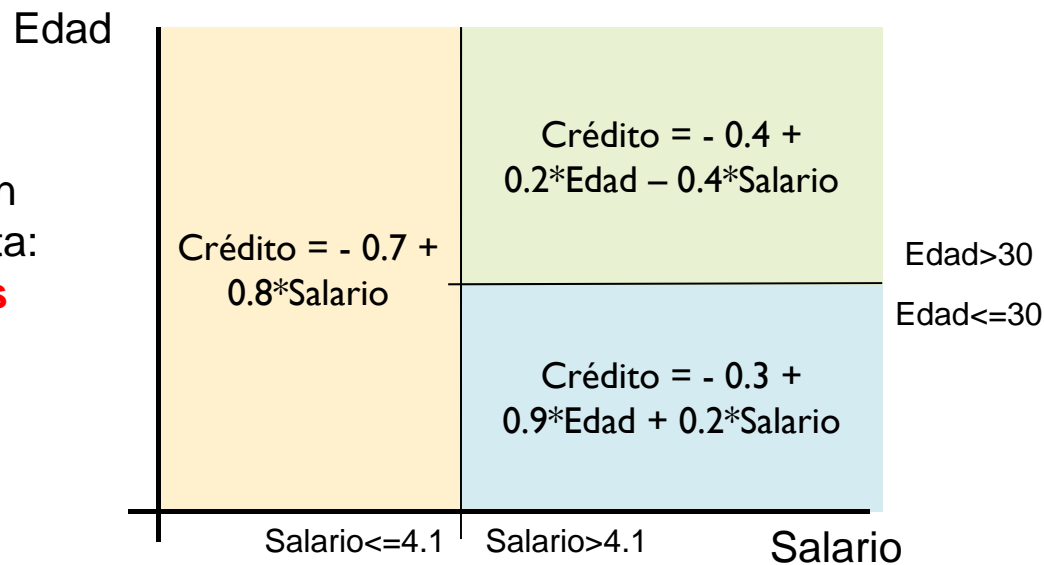
ÁRBOLES DE REGRESIÓN

- Ejemplo:



Cada expresión lineal se aplica en una región distinta:

Modelos locales



ÁRBOLES DE REGRESIÓN

- Ejemplo:

- Base de datos con 209 configuraciones diferentes de ordenadores

	Cycle time (ns) MYCT	Main memory (KB)		Cache (KB) CACH	Channels		Performance PRP
		Min. MMIN	Max. MMAX		Min. CHMIN	Max. CHMAX	
1	125	256	6000	256	16	128	198
2	29	8000	32000	32	8	32	269
3	29	8000	32000	32	8	32	220
4	29	8000	32000	32	8	32	172
5	29	8000	16000	32	8	16	132
...							
207	125	2000	8000	0	2	14	52
208	480	512	8000	32	0	0	67
209	480	1000	4000	0	0	0	45

- Salida: potencia de procesamiento
- Todos los atributos, y la salida, son continuos

ÁRBOLES DE REGRESIÓN

- Ejemplo:

- Árbol de modelos de regresión:

