

Clasificación

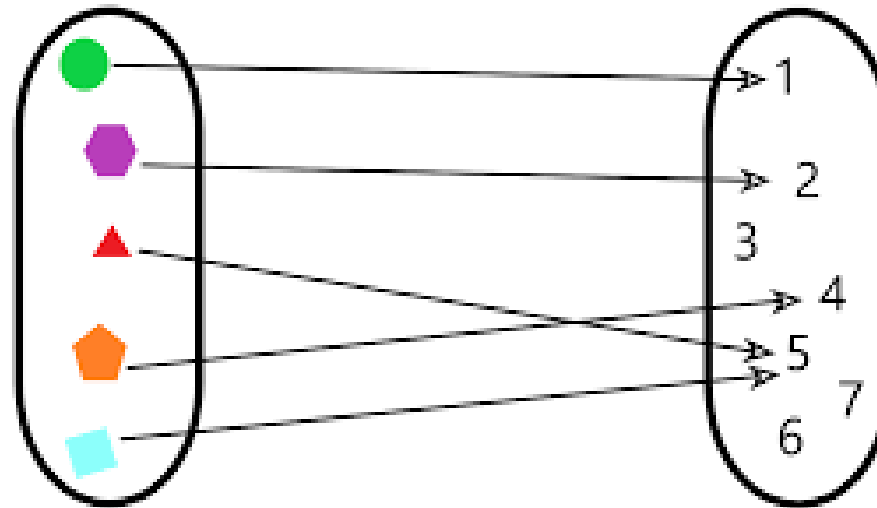
Dra. Consuelo Varinia García Mendoza

Regresión vs clasificación

- En los problemas vistos previamente se crean modelos capaces de predecir valores continuos
 - Precios de casas
 - Clima
 - Valores de acciones
- Para este tipo de problemas los algoritmos de regresión lineal obtienen buenos resultados
- Pero existen otros problemas donde los valores que se desean predecir no son continuos si no categóricos

Clasificación

- Dado un conjunto de ejemplos de entrenamiento compuesto por pares $\{x_i, y_i\}$, encuentre una función $f(x)$ que asigne cada vector de atributos x_i a su clase asociada $y_i, i = 1, 2, \dots, n$, donde n es el número total de ejemplos de entrenamiento.



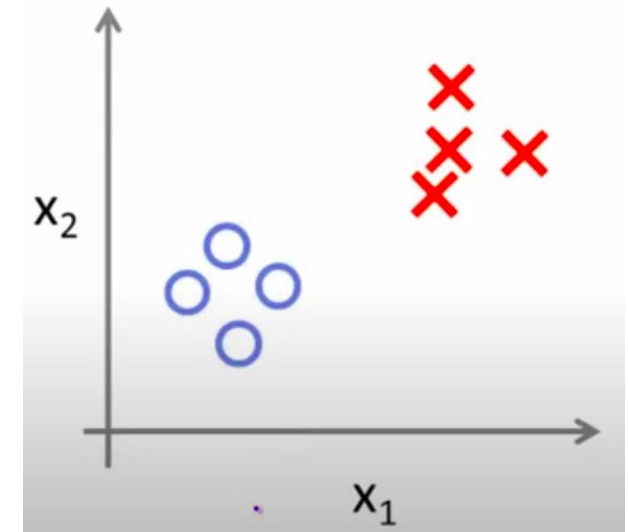
Tipos de clasificación

- Sigle label
 - Binaria
 - Multiclase
- Multi-label
 - Multi-label
 - Multi-output/Multi-target

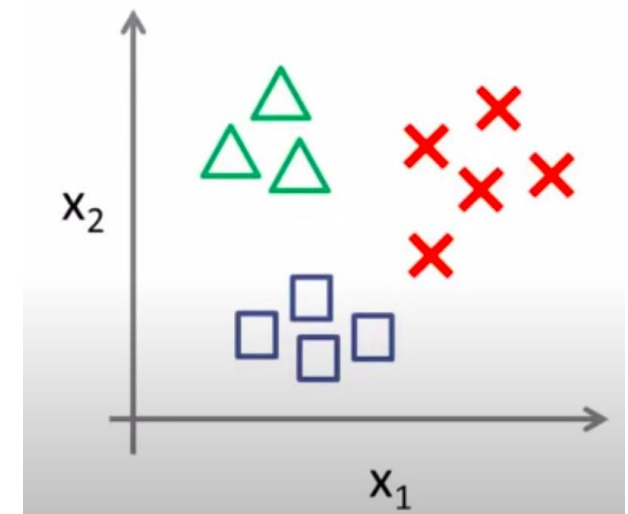
Single-label classification

- **Binaria.** Asocia sólo una clase cada ejemplo o instancia. Una instancia sólo puede pertenecer a la clase 1 o a la clase 2
 - Identificar si un correo es spam o no
 - Identificar si un paciente tiene cáncer o no
- **Multiclase.** Asocia sólo una clase cada ejemplo o instancia. Una instancia puede pertenecer a *la clase 1, clase 2,o clase n* (las clases son mutuamente excluyentes)
 - Polaridad de opinión: [positivo, negativo, neutro]
 - Calificación de un producto (estrellas): [1,2,3,4,5]

Binary classification:



Multi-class classification:



Multi-label classification

- **Multi-label.** Cada instancia puede estar asociada con una o más clases
 - Ejemplo. Identificación de emociones en Tweets de noticias
 - Tweet 1 [alegría, enfado, tristeza, sorpresa, miedo] → [1,0,0,1,0]
 - En el Tweet 1 se presentan las emociones de alegría y sorpresa
 - Clima: [soleado, lluvioso, nublado, nevado]
- **Multi-output / Multi-target.** Cada instancia puede estar asociada con una o más clases y se puede definir la intensidad de pertenencia a cada una de ellas
 - Ejemplo. Identificación de intensidad de emociones en Tweets de noticias
 - Tweet 2 [alegría, enfado, tristeza, sorpresa, miedo] → [0,5,3,0,0]
 - Cada valor de la quintupla puede tomar un valor en el intervalo de 0 a 5 en donde 0 indica que no se presenta presentar la emoción, 1 la intensidad mínima y 5 la intensidad máxima
 - En el Tweet 2 se presentan las emociones de enfado y tristeza con las intensidades 5 y 3 respectivamente

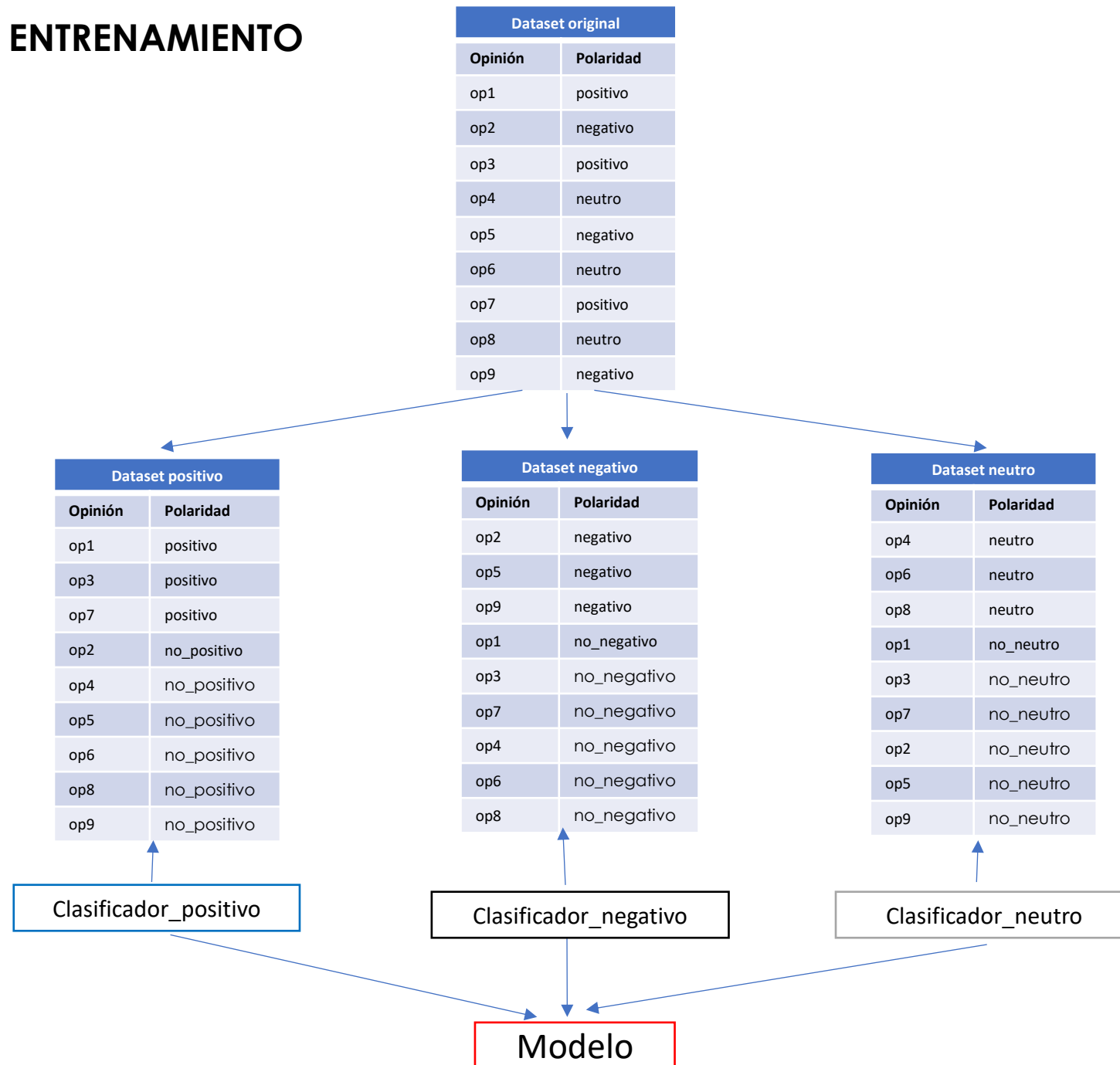
Estrategias de clasificación single-label

- Existen algoritmos de aprendizaje automático que de forma nativa resuelven problemas de clasificación binaria
 - Naïve Bayes
 - Árboles de decisión
 - Redes neuronales
- Para aquellos algoritmos que no manejan la clasificación multiclase de forma nativa existen estrategias que extienden su funcionamiento para poder manejar estos problemas
 - *One vs All* (también llamado *One vs Rest*)
 - *One vs One*

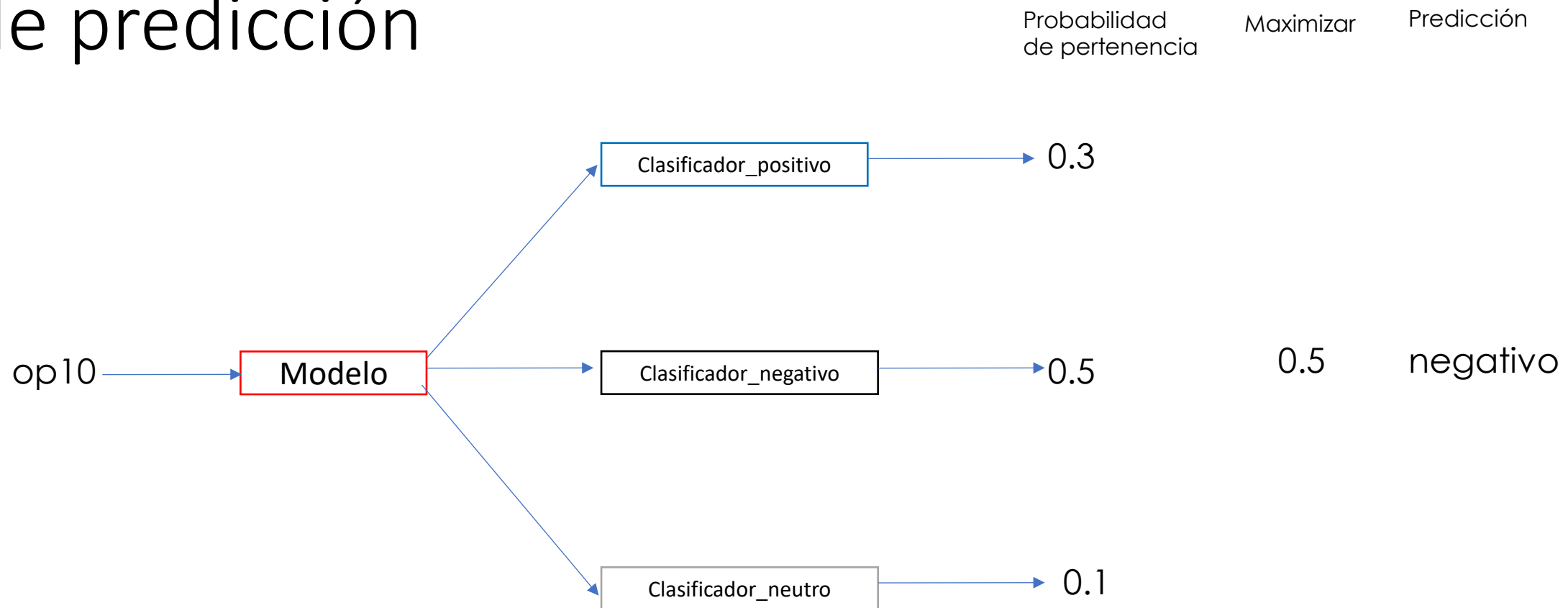
One vs All

- La estrategia parte el problema de clasificación multiclase en múltiples conjuntos de datos de clasificación binaria
- Posteriormente entrena un modelo de clasificación binaria en cada conjunto de datos

ETAPA DE ENTRENAMIENTO



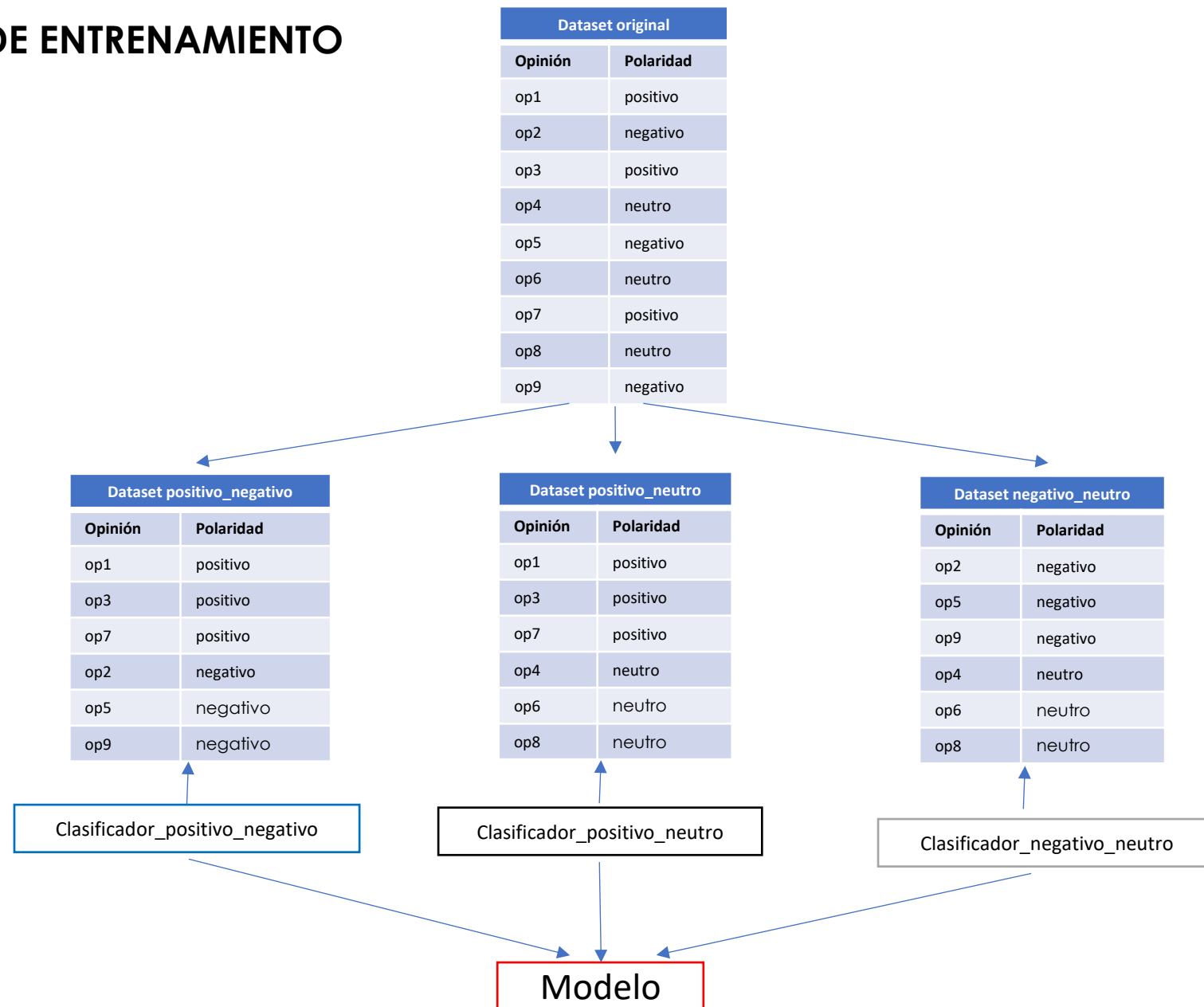
Etapa de predicción



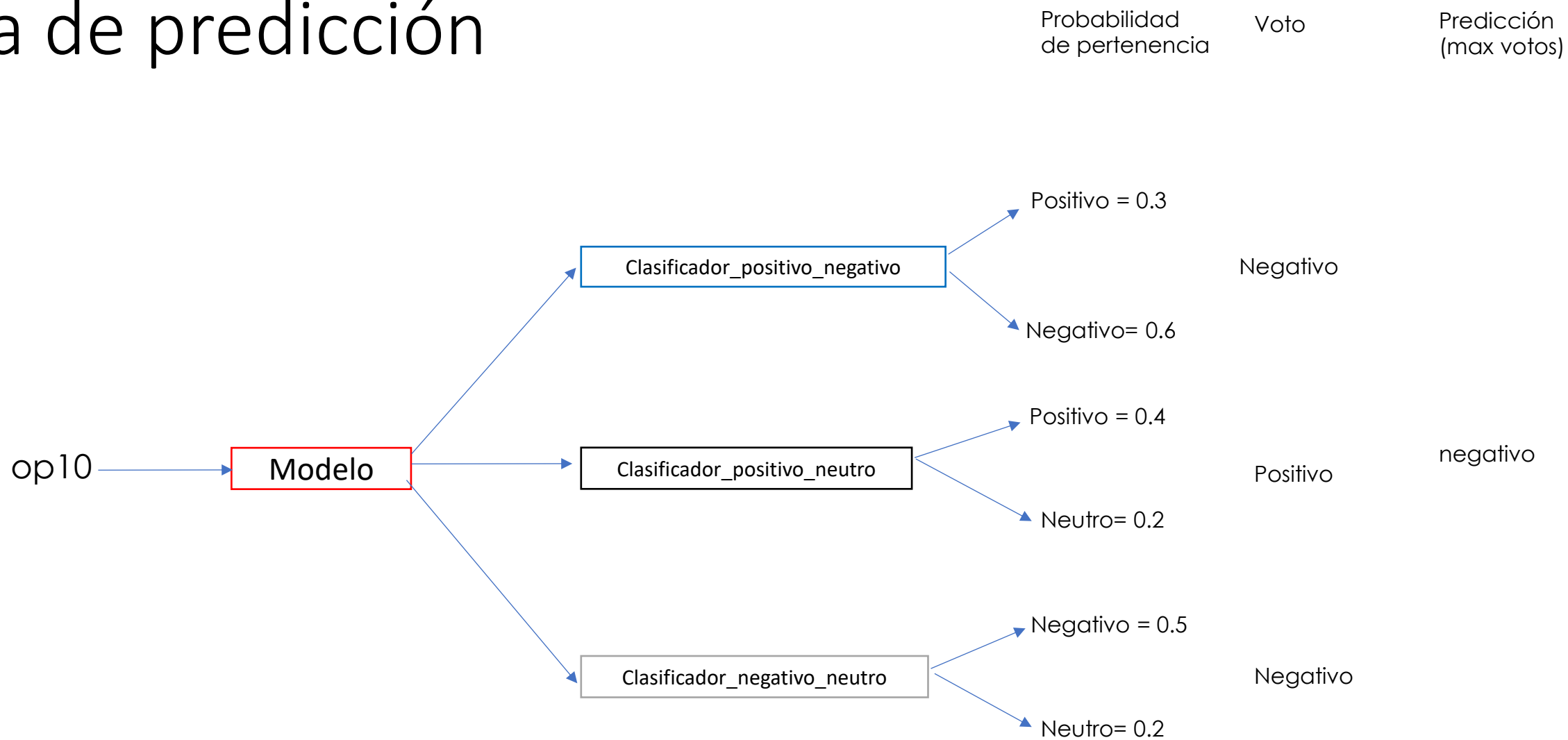
One vs One

- Es una estrategia muy similar a la anterior, pero ahora se crean nuevos datasets formados por todos los pares de clases disponibles

ETAPA DE ENTRENAMIENTO



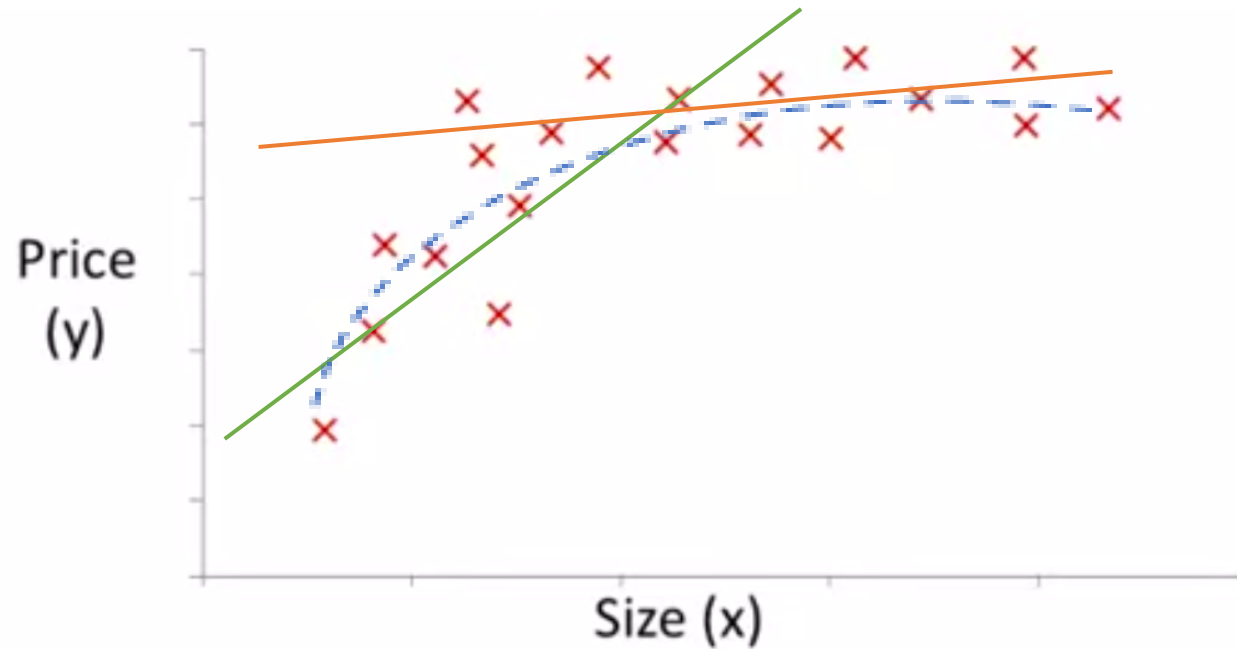
Etapa de predicción



Regresión Logística

Regresión

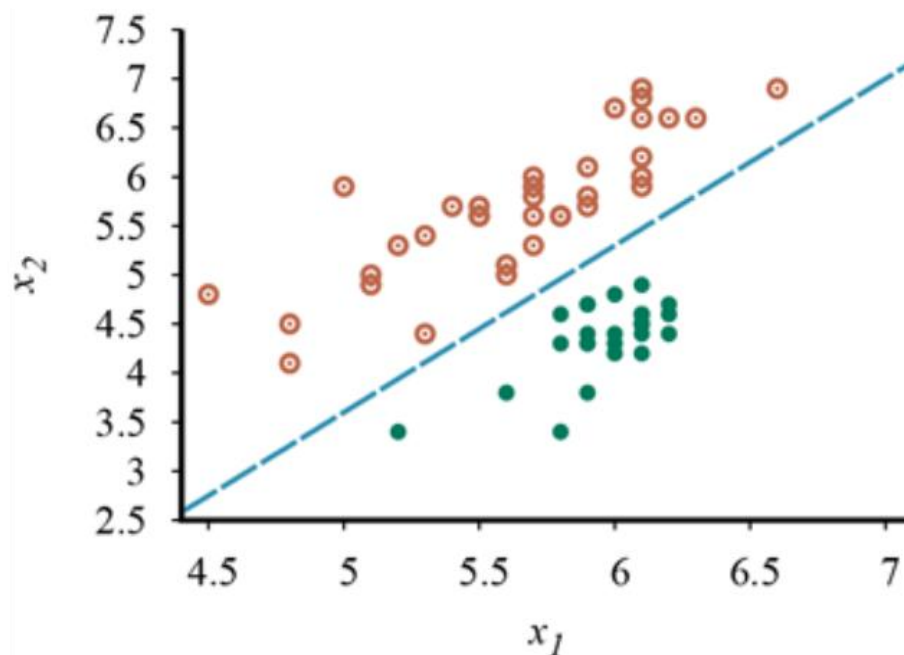
- ¿Se puede utilizar regresión en problemas de clasificación?



Límite de decisión

- Una forma de aplicar la regresión lineal a los problemas de clasificación es estableciendo un límite de decisión (decision boundary)

x_1 : magnitud de la
onda del cuerpo
 x_2 : magnitud de la
onda de superficie



Separador lineal

$$x_2 = 1.7x_1 - 4.9 \quad \text{or} \quad -4.9 + 1.7x_1 - x_2 = 0$$

Explosiones

$$-4.9 + 1.7x_1 - x_2 > 0$$

Terremotos

$$-4.9 + 1.7x_1 - x_2 < 0$$

Simplificando la ecuación

$$-4.9x_0 + 1.7x_1 - x_2 = 0$$

Representación como un vector de pesos

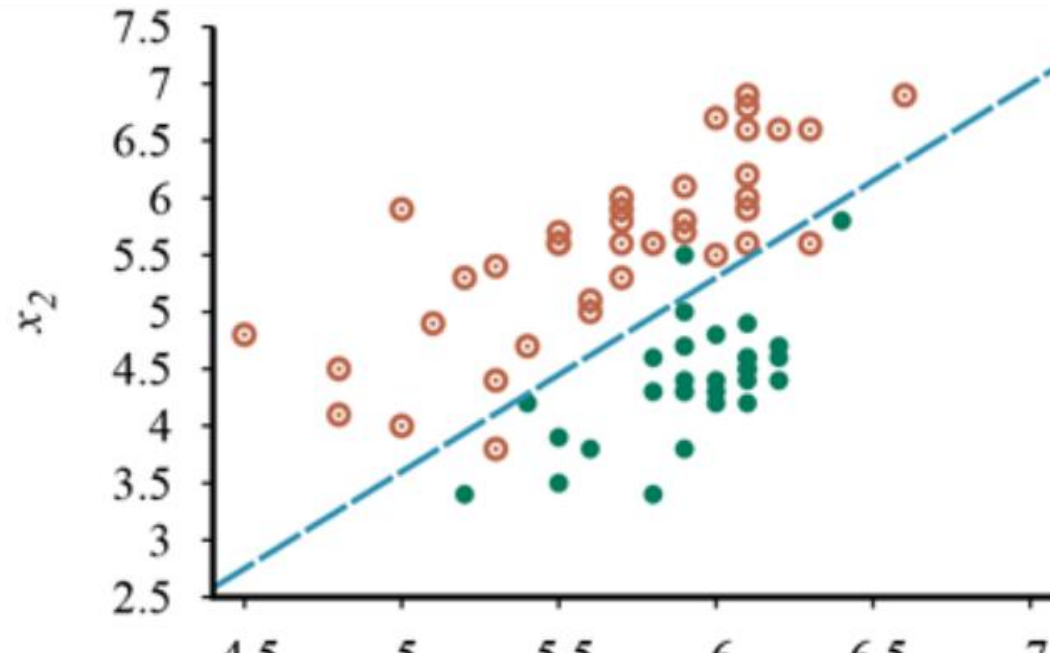
$$\mathbf{w} = \langle -4.9, 1.7, -1 \rangle$$

Hipótesis de clasificación

$$h_{\mathbf{w}}(\mathbf{x}) = 1 \text{ if } \mathbf{w} \cdot \mathbf{x} \geq 0 \text{ and } 0 \text{ otherwise}$$

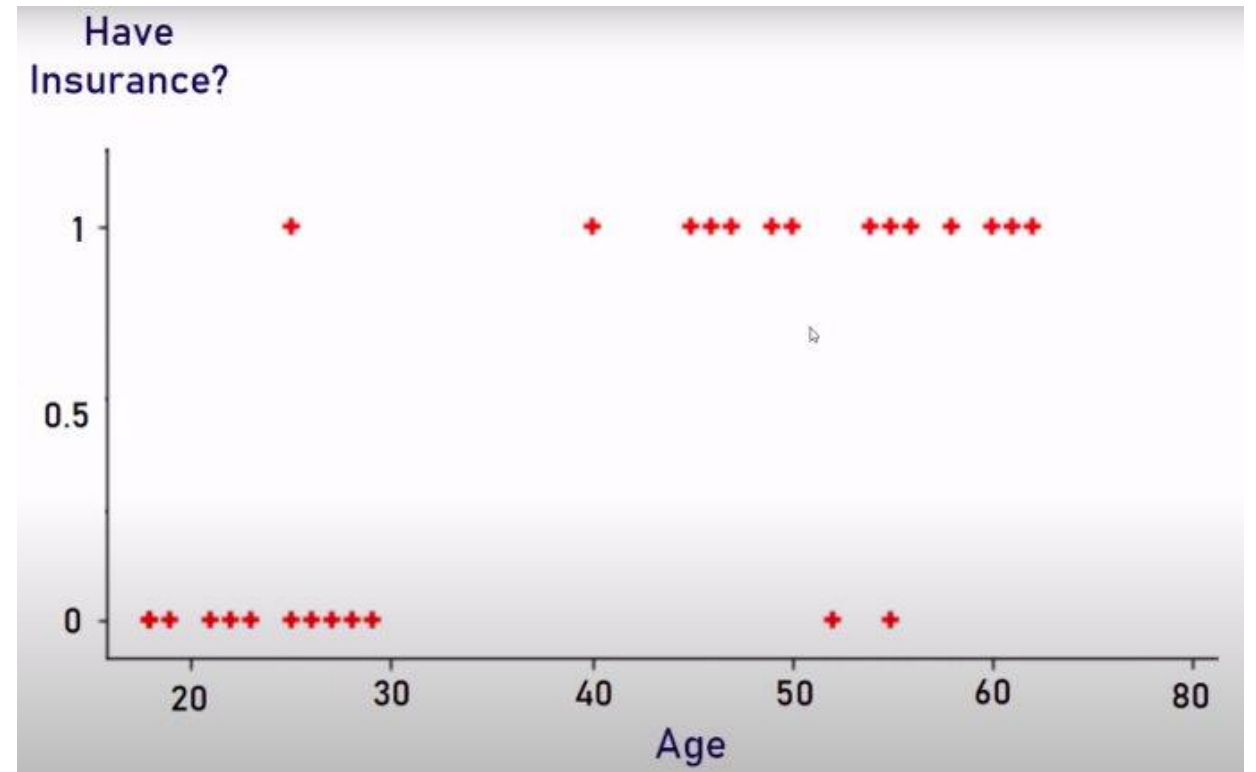
Limitaciones del uso de la regresión lineal para la clasificación

- La regresión lineal asume que los datos son linealmente separables
- Además, el límite de decisión que se establece se ve muy afectado cuando los datos son dispersos



Representación de un problema de clasificación binaria

age	have_insurance
22	0
25	0
47	1
52	0
46	1
56	1
55	0
60	1
62	1
61	1
18	0
28	0
27	0
29	0
49	1

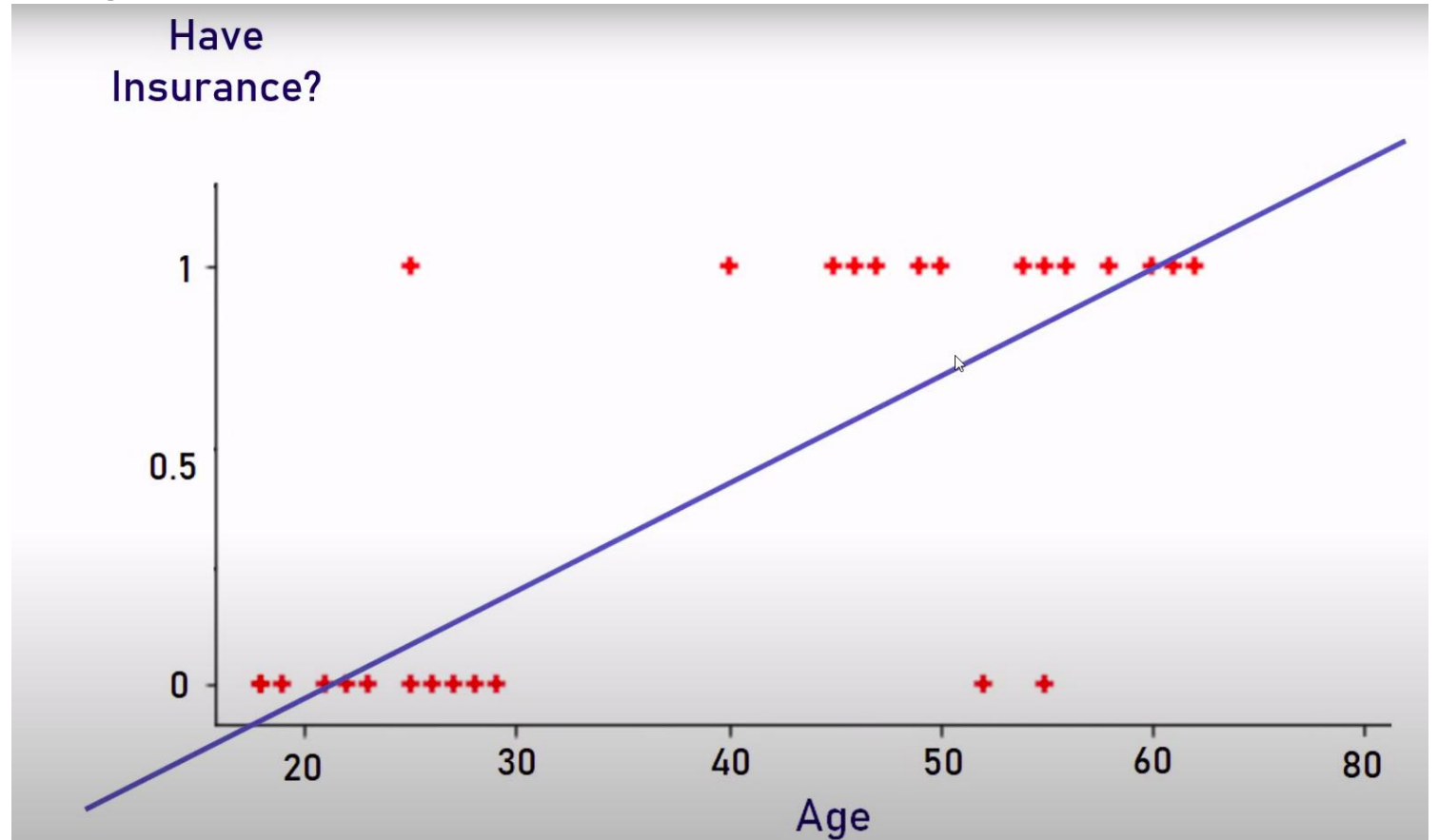


Representación de un problema de clasificación binaria

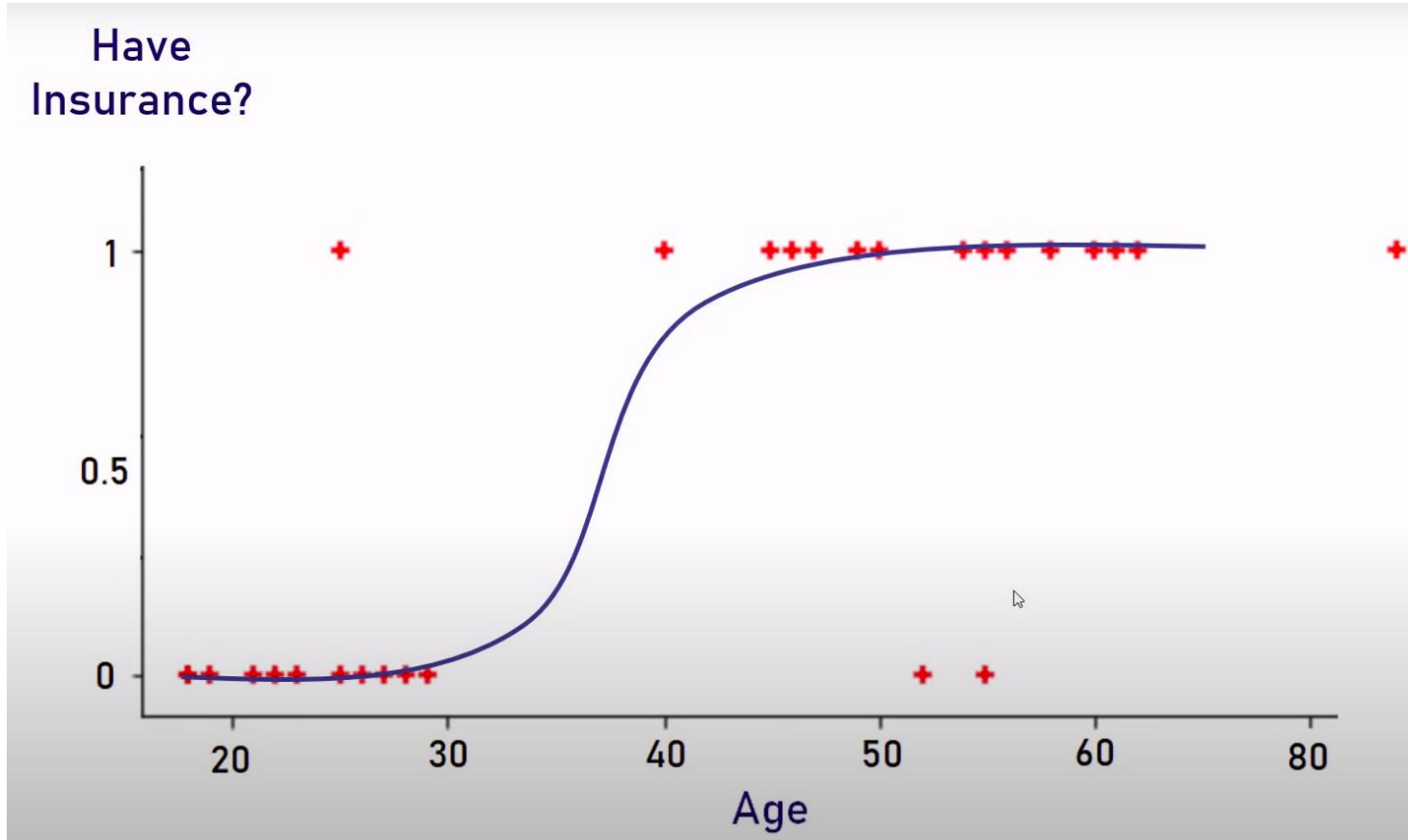
Límite de la decisión

Efecto de datos dispersos

$\hat{y}_i \in \mathbb{R} \in$



Función logística o sigmoidal



- Se restringir la salida $\hat{y}_i \in [0, 1]$, en lugar de $\hat{y}_i \in \mathbb{R}$ como sucede en la regresión lineal
- Los valores menores a 0.5 pertenecen a la clase 0 y los mayores a 0.5 a la clase 1

Función sigmoide

- La ecuación de la función sigmoide se define de la siguiente manera:

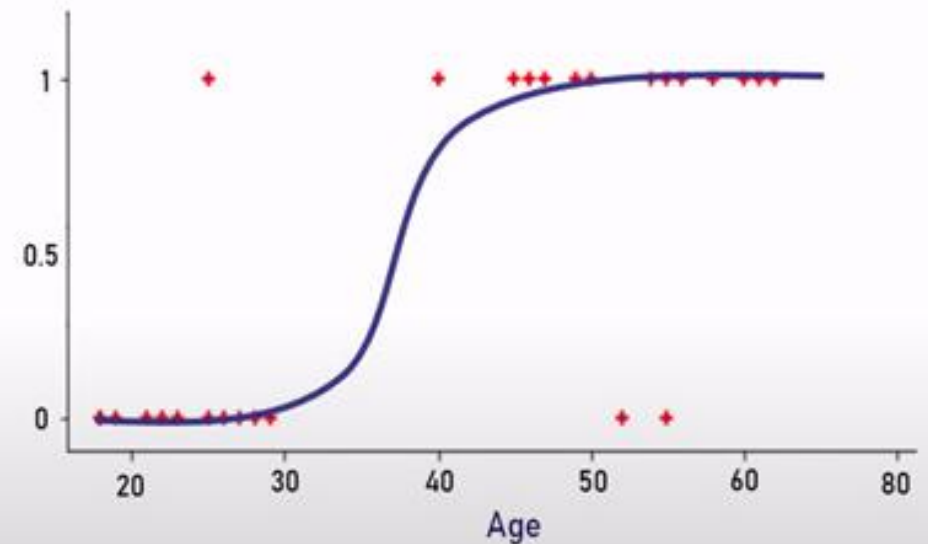
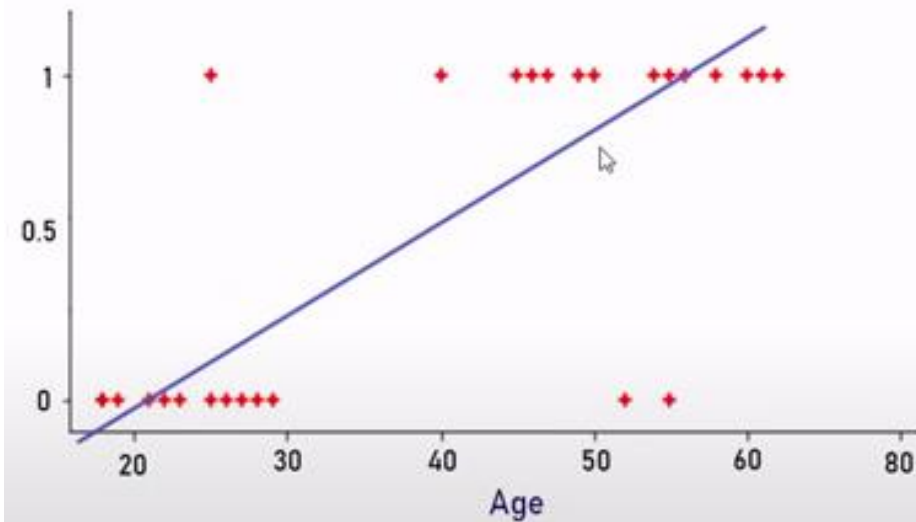
$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

e = número de Euler ~ 2.71828

Hipótesis de la regresión logística

$$z = w_1x + w_0$$

$$\sigma(z) = \frac{1}{1 + e^{-(w_1x + w_0)}}$$



Regresión logística

$$p(y = 1|x) = \sigma(z)$$

$$p(y = 0|x) = 1 - \sigma(z) = \sigma(-z)$$

donde

$$z = w_1x + w_0$$

$$\sigma(z) = \frac{1}{1 + e^{-(w_1x + w_0)}}$$

Ejemplo. Reseñas de películas

kaggle

+

Create

Home

Competitions

Datasets

Models

Code

Discussions

Learn

More

Your Work

VIEWED

IMDB Dataset of 50...

Search

LUIS DIEGO FERNANDEZ · UPDATED 3 YEARS AGO

19

New Notebook

Download (55 MB)

IMDb

IMDB Dataset of 50K Movie Reviews (Spanish)

50K Movie Reviews Translated to Spanish/50k Reseñas de Cine Traducidas a Español

Data Card

Code (3)

Discussion (0)

Suggestions (0)

About Dataset

Context / Contexto

English

This dataset was created using the "IMDB Dataset of 50K Movie Reviews", which contains 50k movie reviews and a positive or negative sentiment attach to them. This dataset contains the same data with the addition of the spanish translation of each review. The data was translated using the googletrans python library, which itself uses the google

Usability ⓘ

10.00

License

Other (specified in description)

Expected update frequency

Quarterly

Reseña de prueba $k + 1$

It's hokey. There are virtually no surprises, and the writing is second-rate. So why was it so enjoyable? For one thing, the cast is great. Another nice touch is the music. I was overcome with the urge to get off the couch and start dancing. It sucked me in, and it'll do the same to you.

Var	Definition	Value
x_1	count(positive lexicon words \in doc)	
x_2	count(negative lexicon words \in doc)	
x_3	$\begin{cases} 1 & \text{if "no"} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	
x_4	count(1st and 2nd pronouns \in doc)	
x_5	$\begin{cases} 1 & \text{if "!"} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	
x_6	ln(word count of doc)	

It's **hokey**. There are virtually **no** surprises, and the writing is **second-rate**. So why was it so **enjoyable**? For one thing, the cast is **great**. Another **nice** touch is the music. **I** was overcome with the urge to get off the couch and start dancing. It sucked **me** in, and it'll do the same to **you**.

Var	Definition	Value
x_1	count(positive lexicon words \in doc)	3
x_2	count(negative lexicon words \in doc)	2
x_3	$\begin{cases} 1 & \text{if "no"} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	1
x_4	count(1st and 2nd pronouns \in doc)	3
x_5	$\begin{cases} 1 & \text{if "!"} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	0
x_6	$\ln(\text{word count of doc})$	$\ln(66) = 4.19$

Supongamos que ya pasamos por la etapa de entrenamiento de regresión logística y se aprendió el vector de pesos $w = [2.5 \quad -5 \quad -1.2 \quad 0.5 \quad 2.0 \quad 0.7]$ y $b = 0.1$. Podemos calcular la probabilidad de que el texto de prueba sea una crítica positiva $p(y = 1|x)$ y la probabilidad de que sea una crítica negativa $p(y = 0|x)$

$$p(+|x) = P(y = 1|x) = \sigma(\mathbf{w} \cdot \mathbf{x} + \mathbf{b}) = \sigma\left(\begin{bmatrix} 2.5 \\ -5 \\ -1.2 \\ 0.5 \\ 2 \\ 0.7 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \\ 1 \\ 3 \\ 0 \\ 4.19 \end{bmatrix} + 0.1\right) = 0.7$$

$$p(-|x) = P(y = 0|x) = 1 - \sigma(\mathbf{w} \cdot \mathbf{x} + \mathbf{b}) = 0.3$$

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Por lo tanto la reseña es una reseña positiva

e = número de Euler ~ 2.71828

$$z = \mathbf{w} \cdot \mathbf{x} + b$$

Regresión logística (LR)

- LR, así como Naïve Bayes es un clasificador probabilístico de AA
- LR es una de las herramientas analíticas más importantes de las ciencias sociales y naturales
- En el PLN, la LR es el algoritmo de AA supervisado de referencia para la clasificación, y también tiene una relación muy estrecha con las redes neuronales
- Una red neuronal puede verse como una serie de clasificadores de LR apilados unos sobre otros
- La entropía cruzada(función de pérdida de la LR) es convenientemente convexa, es decir sólo tiene un mínimo; no hay mínimos locales, por lo que el descenso del gradiente a partir de cualquier punto tiene garantizado el éxito.