



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# APRENDIZAJE DE MÁQUINAS

## Métodos Supervisados

**JOHN W. BRANCH**

Profesor Titular

Departamento de Ciencias de la Computación y de la Decisión Director del  
Grupo de I+D en Inteligencia Artificial – GIDIA  
[jwbranch@unal.edu.co](mailto:jwbranch@unal.edu.co)

*Tomado y actualizado de presentaciones del docente:*

**Carlos Madrigal**

Departamento de Ciencias de la Computación y de la Decisión Investigador del Grupo de I+D en  
Inteligencia Artificial – GIDIA

*Gidia*  
Grupo de I+D  
en Inteligencia Artificial



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# Contenido

## 1. Regresión

- a. Regresión Lineal
- b. Regresión Polinomial
- c. Regresión Lasso, Ridge y Elastic-net
- d. Árbol de Decisión para Regresión
- e. Red Neuronal para Regresión

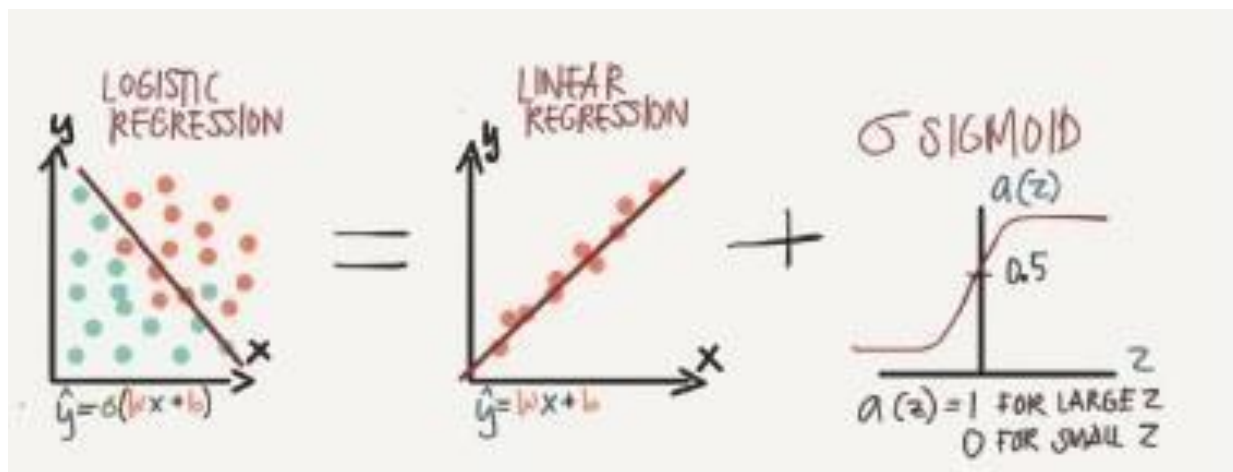
## 2. Clasificación

- a. Regresión Logística
- b. K-nn
- c. Máquina de Vectores de Soporte
- d. Algoritmo XGBoost para Clasificación
- e. Red Neuronal para Clasificación

# Regresión Logística

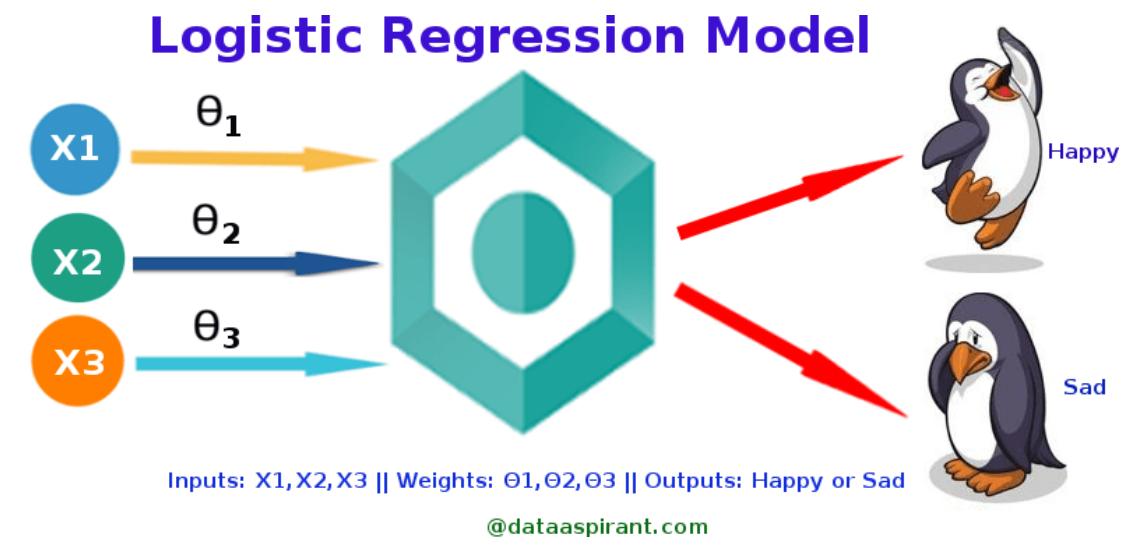
## DEFINICIÓN:

La regresión logística es un modelo de clasificación que se utiliza para **predecir la probabilidad**  $P(y = 1)$  **de una variable dependiente categórica** en función de  $x$ . Así, la variable  $y$  es una variable binaria codificada como 1 (positivo, éxito, etc.) o 0 (negativo, falla, etc.).



## Algunos ejemplos de aplicación:

- E-mail: spam/no spam
- Transacciones en línea: fraude/no fraude
- Tumores: maligno/no maligno



# Regresión Logística

## VENTAJAS:

- Es un modelo de clasificación eficiente y simple.
- No es necesario disponer de grandes recursos computacionales.
- Los resultados son altamente interpretables.



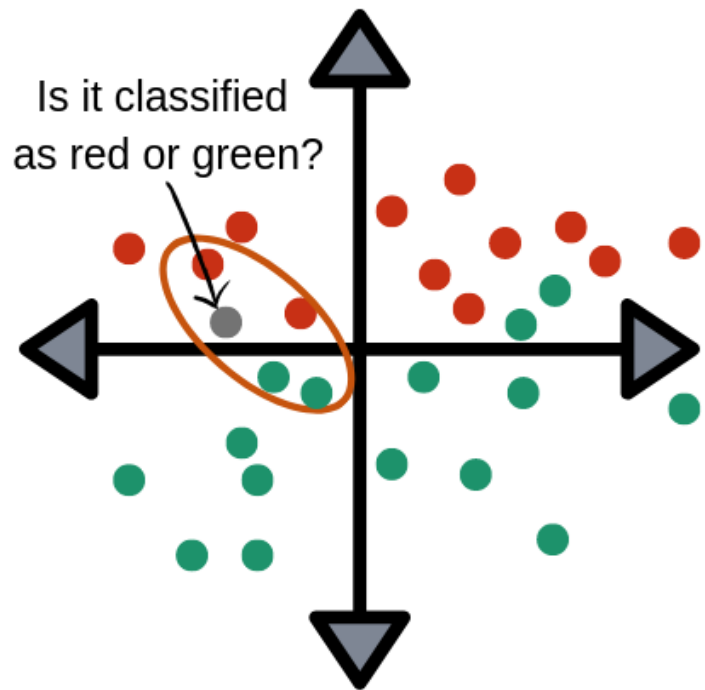
## DESVENTAJAS:

- Imposibilidad de resolver directamente problemas no lineales.
- La variable objetivo esta ha de ser linealmente separable.
- La regresión logística no es uno de los algoritmos más potentes que existen.



# Clasificador KNN

La idea básica sobre la que se fundamenta este paradigma es que un nuevo caso se va a clasificar en la clase más frecuente a la que pertenecen sus  $K$  vecinos más cercanos



## COMIENZO

Entrada:  $D = \{(\mathbf{x}_1, c_1), \dots, (\mathbf{x}_N, c_N)\}$

$\mathbf{x} = (x_1, \dots, x_n)$  nuevo caso a clasificar

PARA todo objeto ya clasificado  $(x_i, c_i)$

calcular  $d_i = d(\mathbf{x}_i, \mathbf{x})$

Ordenar  $d_i (i = 1, \dots, N)$  en orden ascendente

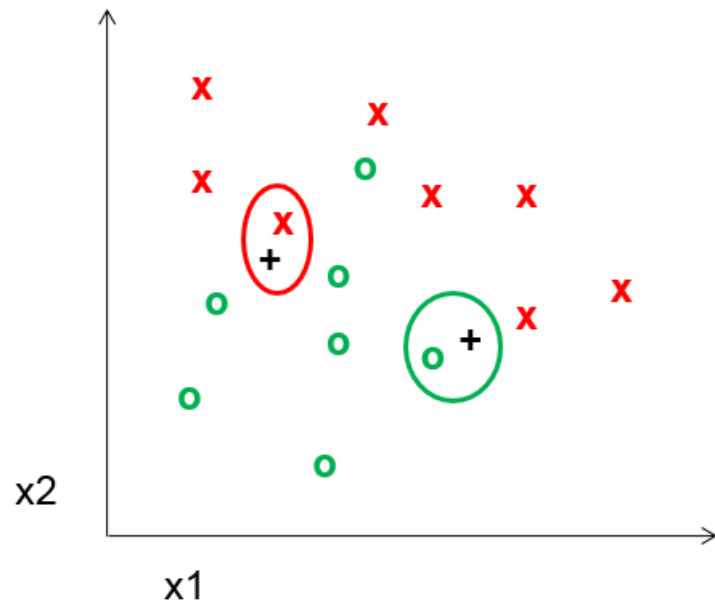
Quedarnos con los  $K$  casos  $D_{\mathbf{x}}^K$  ya clasificados más cercanos a  $\mathbf{x}$

Asignar a  $\mathbf{x}$  la clase más frecuente en  $D_{\mathbf{x}}^K$

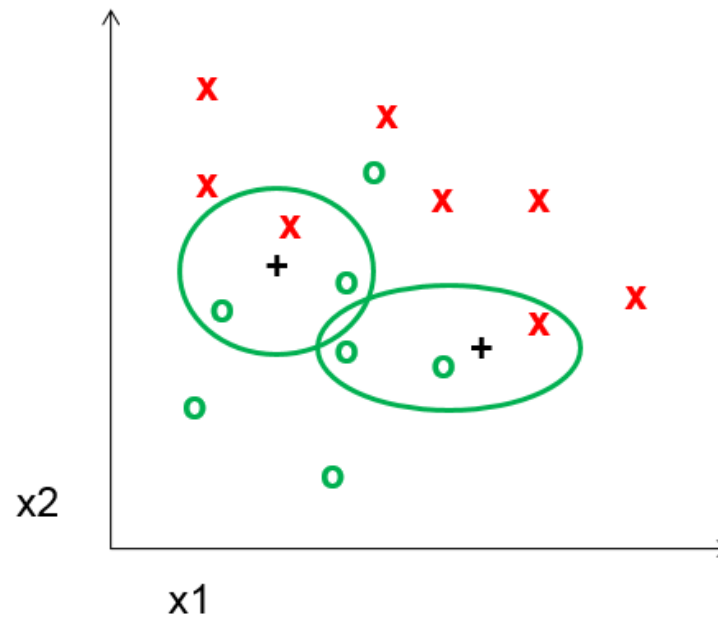
FIN

# Clasificador KNN

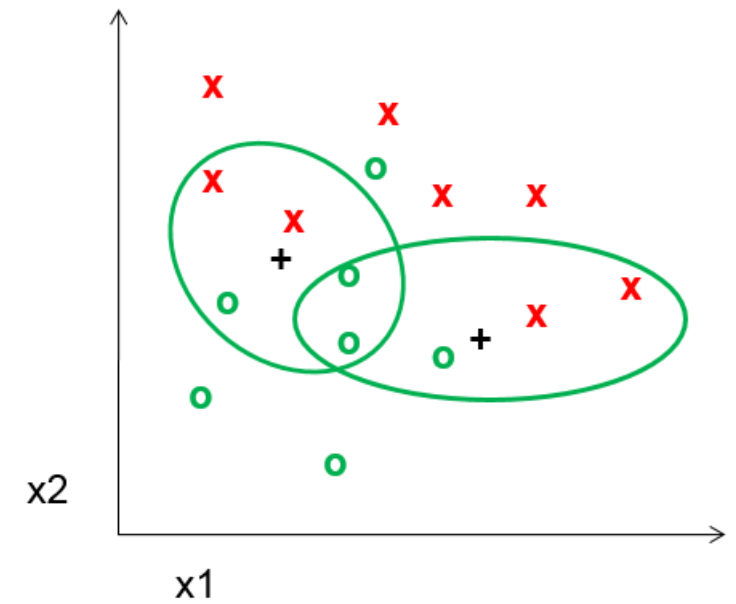
KNN-1 (1 Vecino más cercano)



KNN-3 (3 Vecinos más cercanos)



KNN-5 (5 Vecinos más cercanos)

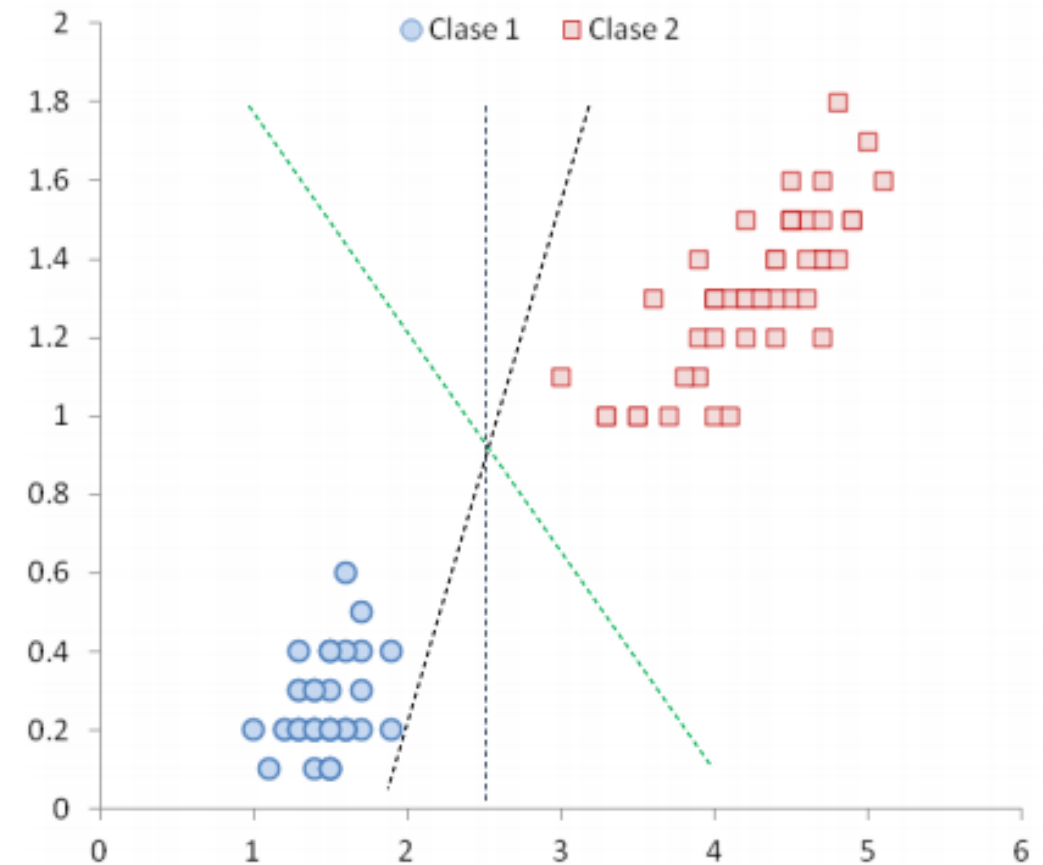


# Máquinas de Vectores de Soporte

Las Máquinas de Vectores de Soporte (SVM) buscan el hiperplano que separe de forma óptima los puntos de una clase de otra.

En un problema de dos clases, linealmente separables, pueden existir muchas fronteras de decisión (o hiperplanos) que pueden separar las clases.

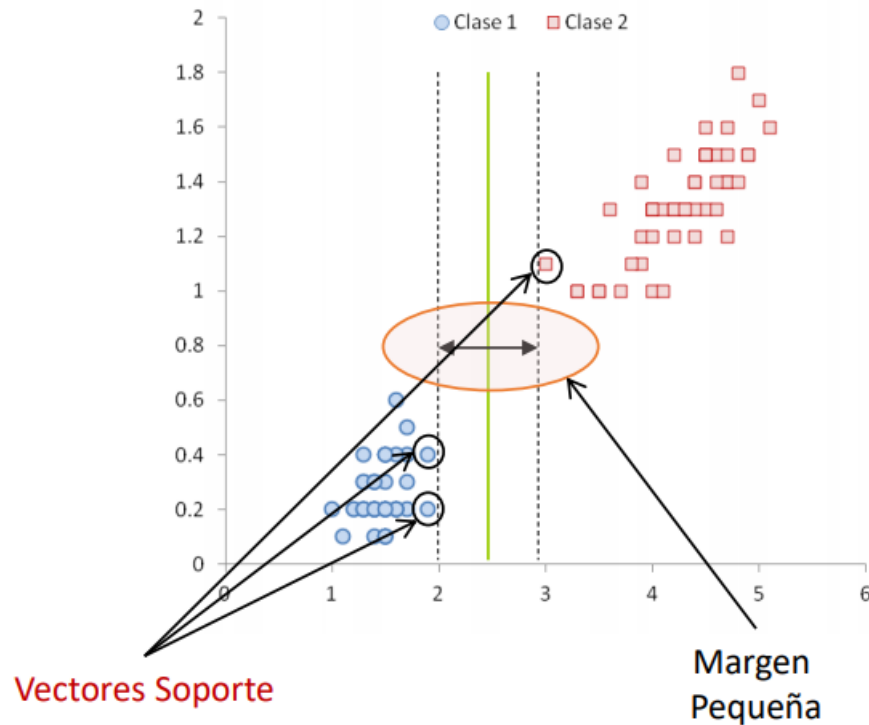
Sin embargo, ¿Son todas esas fronteras igual de buenas?



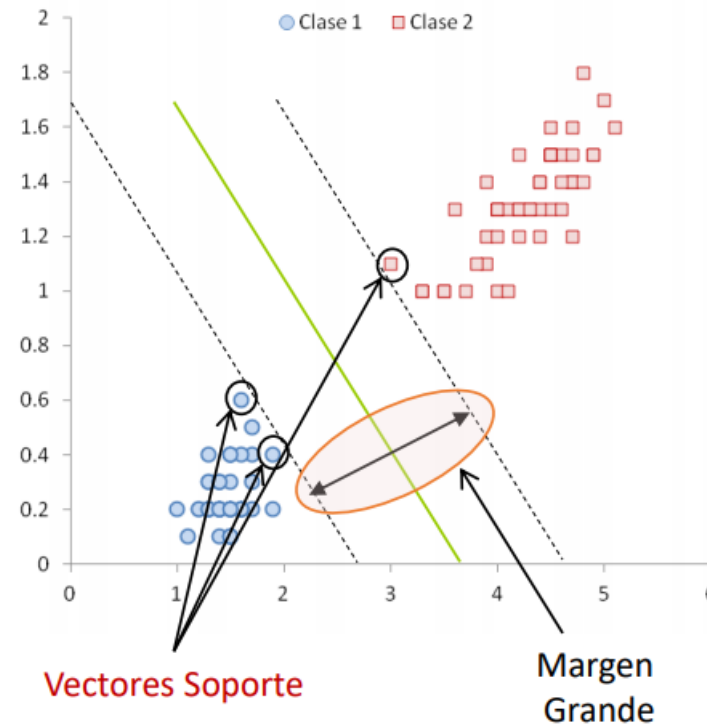


# Máquinas de Vectores de Soporte

La SVM busca el hiperplano que maximiza la distancia (o margen) con los puntos que estén más cerca de él, razón por la cual también se les conoce a las SVM como clasificadores de margen máximo.



Como el hiperplano separa las muestras positivas (+1) de las negativas (-1), los puntos que están en el hiperplano deben satisfacer la ecuación:  $\mathbf{w}^T \mathbf{x} + b = 0$ .



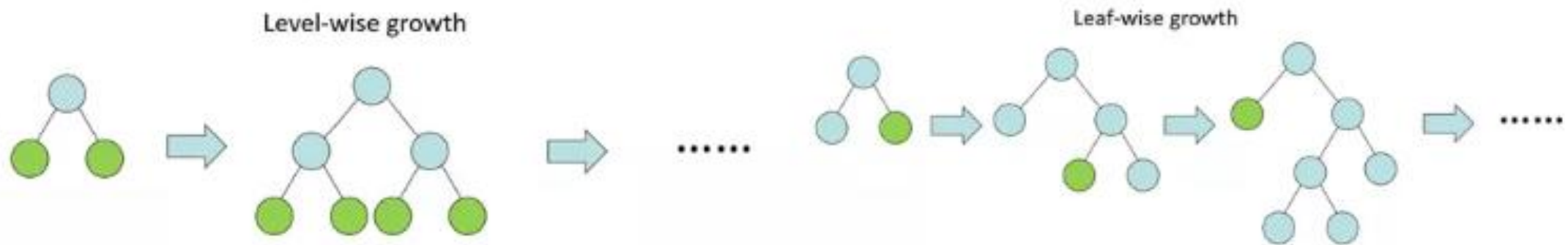
# XGBoost: eXtreme Gradient Boosting

Links de interés:

- <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>

Xgboost es un algoritmo de ensamble, tipo boosting, de árboles de decisión. En un ensamble boosting los árboles son contruidos de manera secuencial, por lo que cada árbol siguiente reduce los errores de los árboles previos. Cada árbol aprende de sus predecesores y actualiza los errores.

Xgboost es uno de los métodos más populares de modelado de bases de datos tabulares de cualquier tamaño, es muy rápido y escalable.



*An illustration demonstrating the difference between level-wise and leaf-wise growth*

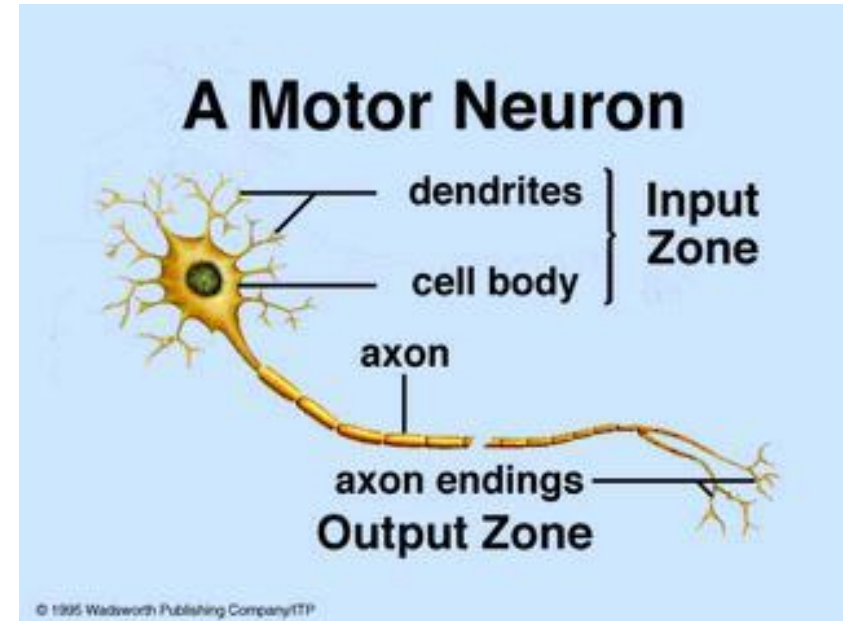
# Red Neuronal para Clasificación

**Cuerpo:** Soma celular

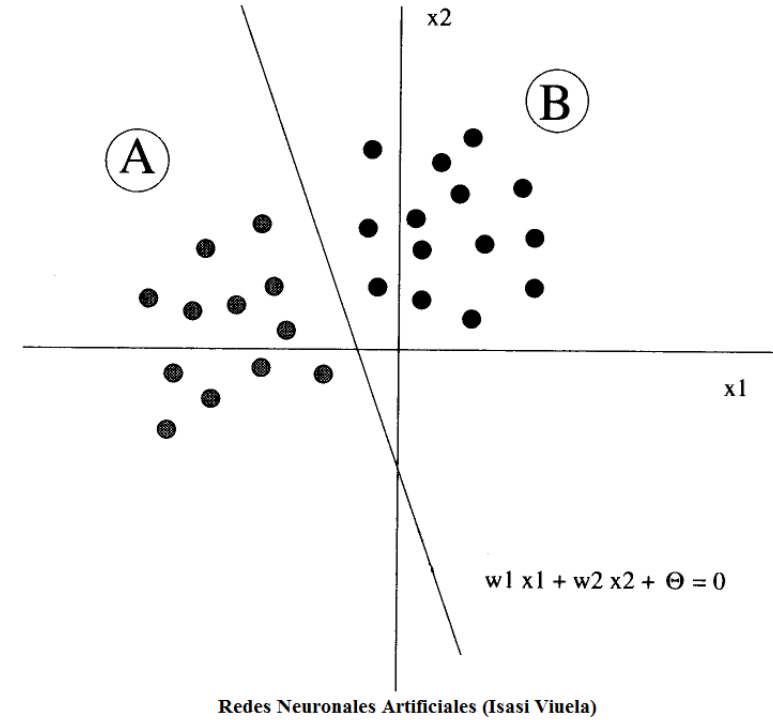
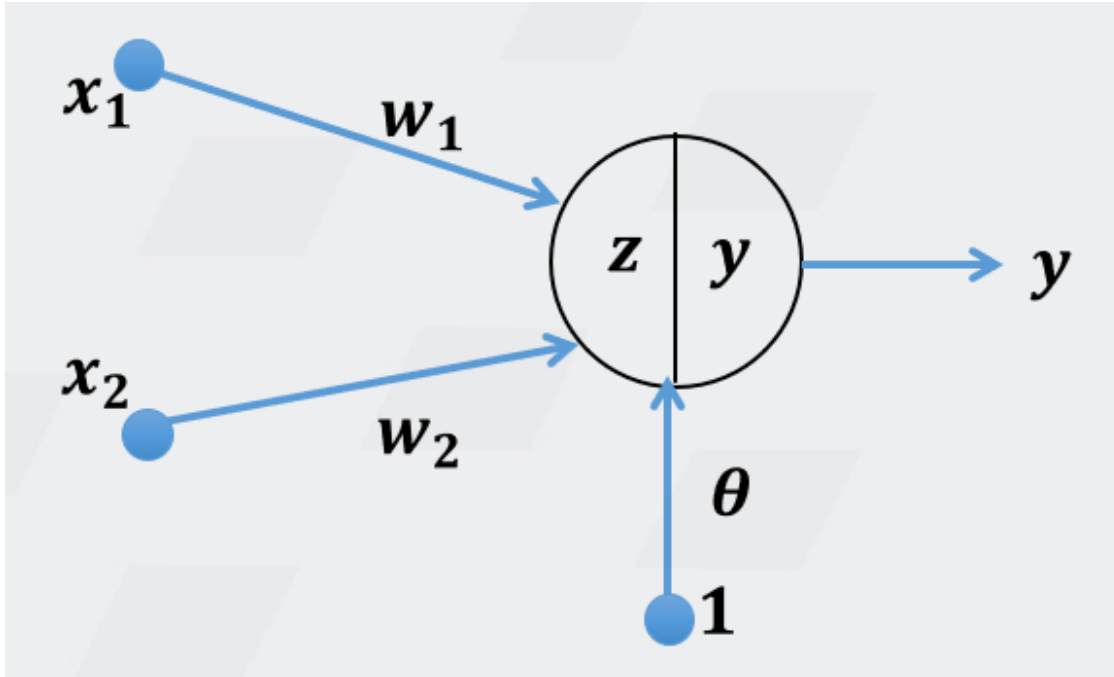
**Axon:** Ramificación de la célula

**Dendritas:** Ramificación de entrada

**Sinapsis:** Recoge información electro-química



# Red Neuronal para Clasificación



Redes Neuronales Artificiales (Isasi Viuela)

# Red Neuronal para Clasificación

*“El objetivo del entrenamiento de una RNA es conseguir que una aplicación determinada, para un conjunto de entradas produzca el conjunto de salidas deseadas o mínimamente consistentes”*

***El conjunto de Aprendizaje debe ser:***

*Significativo: Número suficiente de ejemplos.*

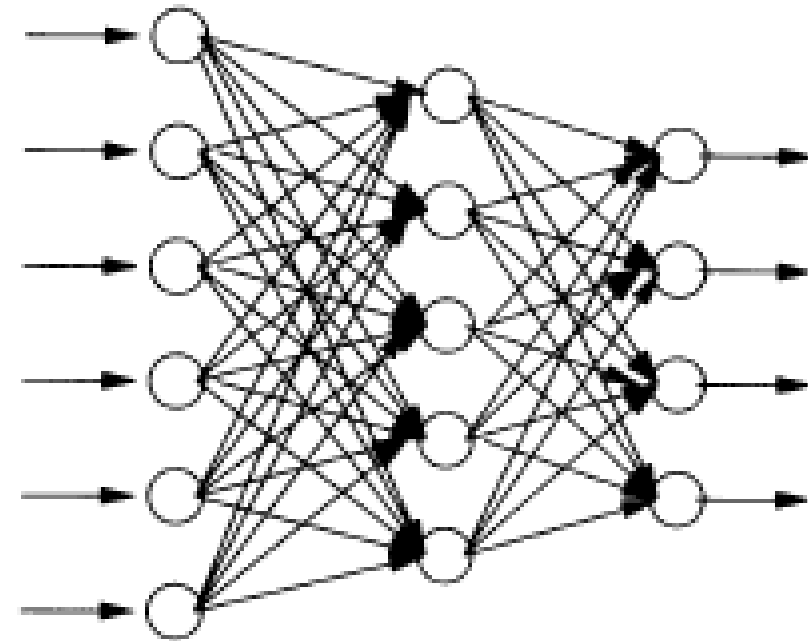
*Representativo: Conjunto de aprendizaje diverso.*

***La convergencia en el entrenamiento se da por:***

*Número fijo de ciclos.*

*Error descienda de cierta cantidad.*

*Modificación de los pesos sea irrelevante.*





# Reconocimiento de Patrones

## Ejemplo Práctico - Mandarinas vs. Naranjas

[ Capítulo 1 ]

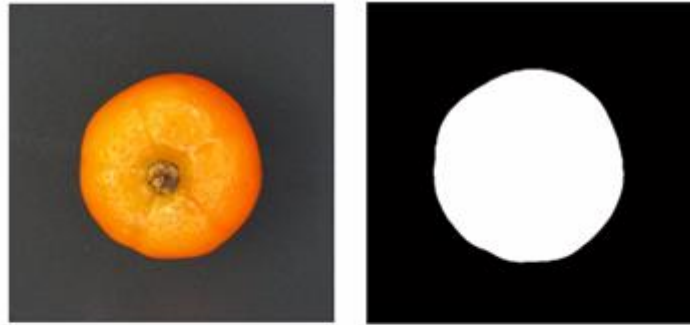
Domingo Mery

Departamento de Ciencia de la Computación  
Escuela de Ingeniería  
Universidad Católica de Chile

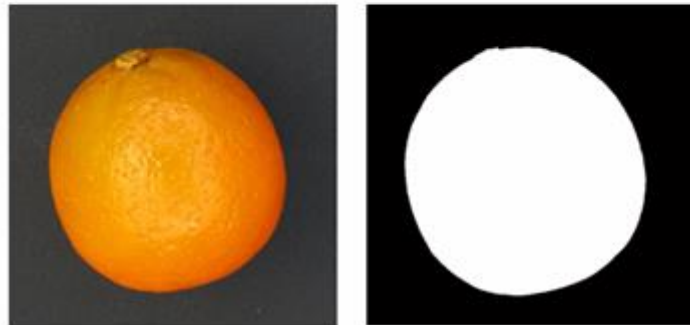


¿cómo separar las mandarinas de las naranjas?

Medición del tamaño es una buena alternativa: (*las mandarinas son más pequeñas*)



Área = 15.457 pixeles



Área = 18.583 pixeles

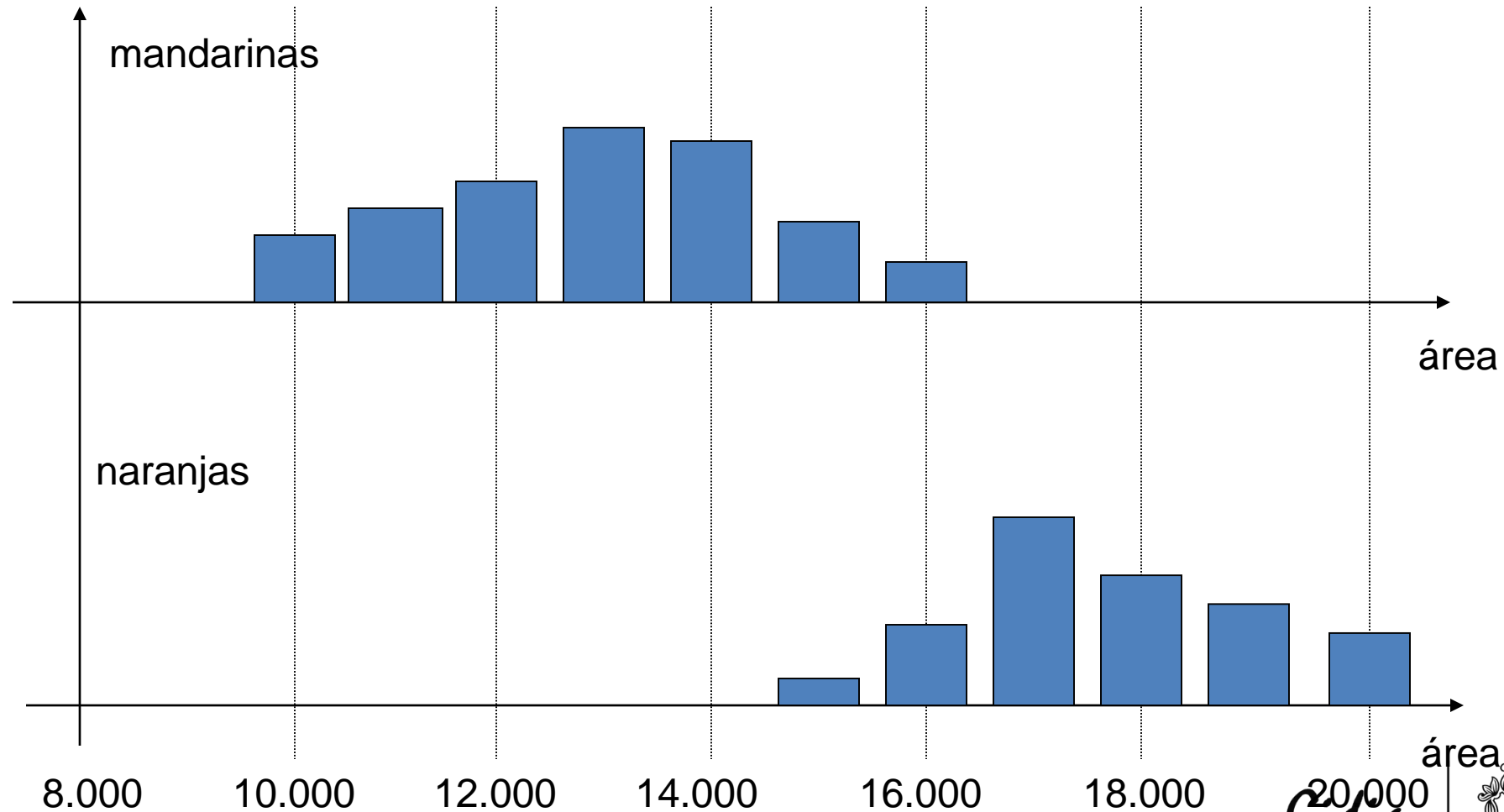


Recolección de información: Área en Pixeles

Naranja-01	19.327	Mandarina-01	13.221
Naranja-02	18.265	Mandarina-02	14.987
Naranja-03	17.456	Mandarina-03	15.321
Naranja-04	19.341	Mandarina-04	15.987
Naranja-05	16.342	Mandarina-05	16.345
Naranja-06	16.987	Mandarina-06	15.965
Naranja-07	17.001	Mandarina-07	16.341
:	19.056	:	
Naranja-75	15.900	Mandarina-50	13.439

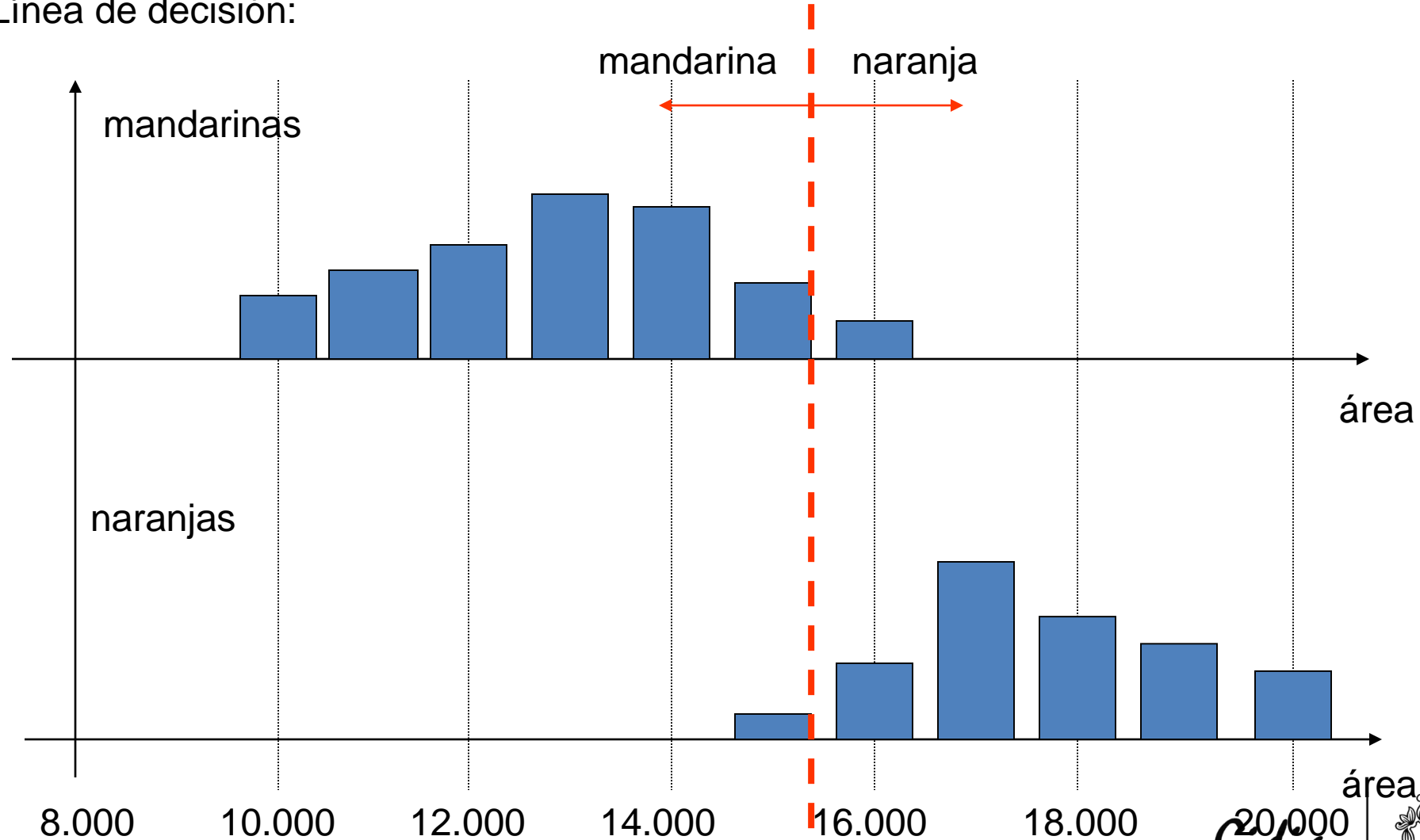
# Un ejemplo práctico

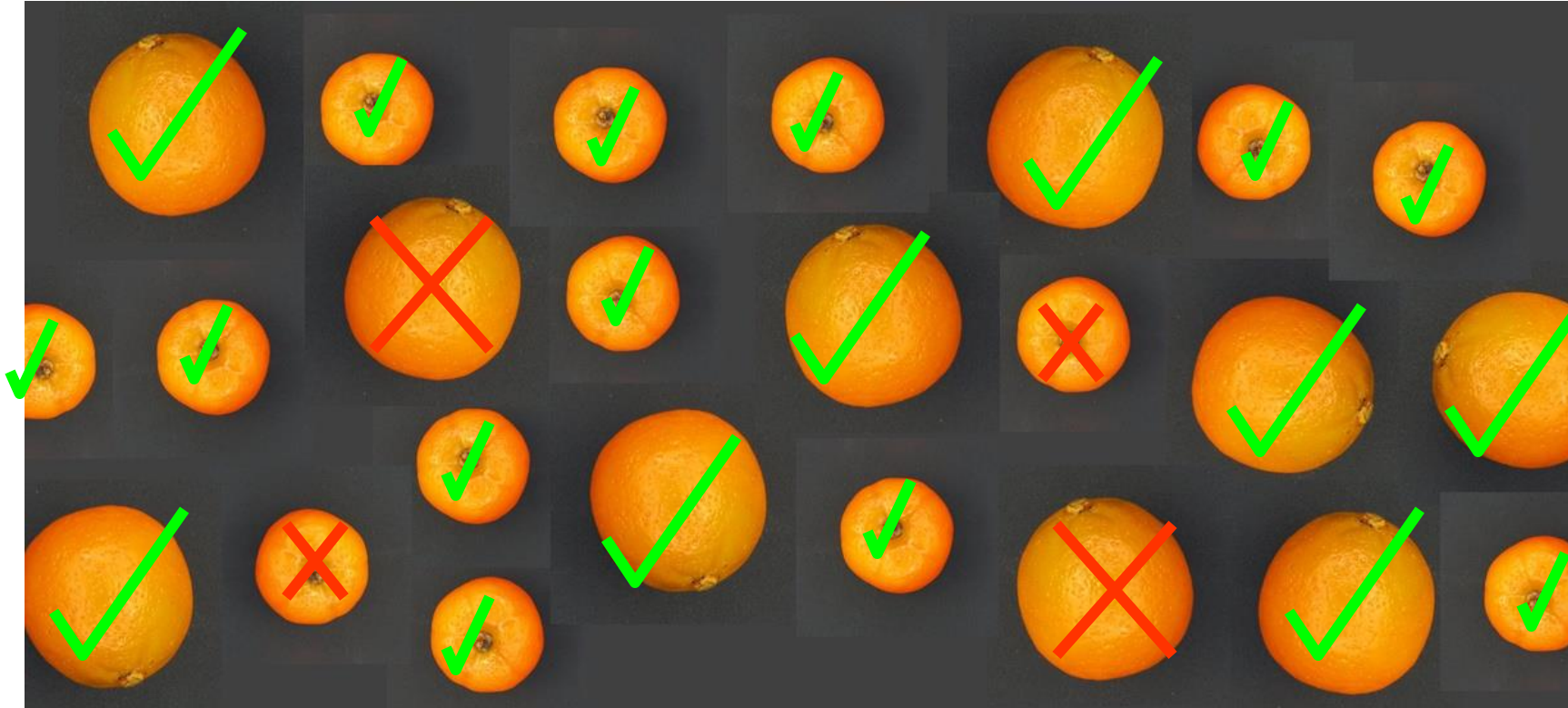
Histogramas:



## Un ejemplo práctico

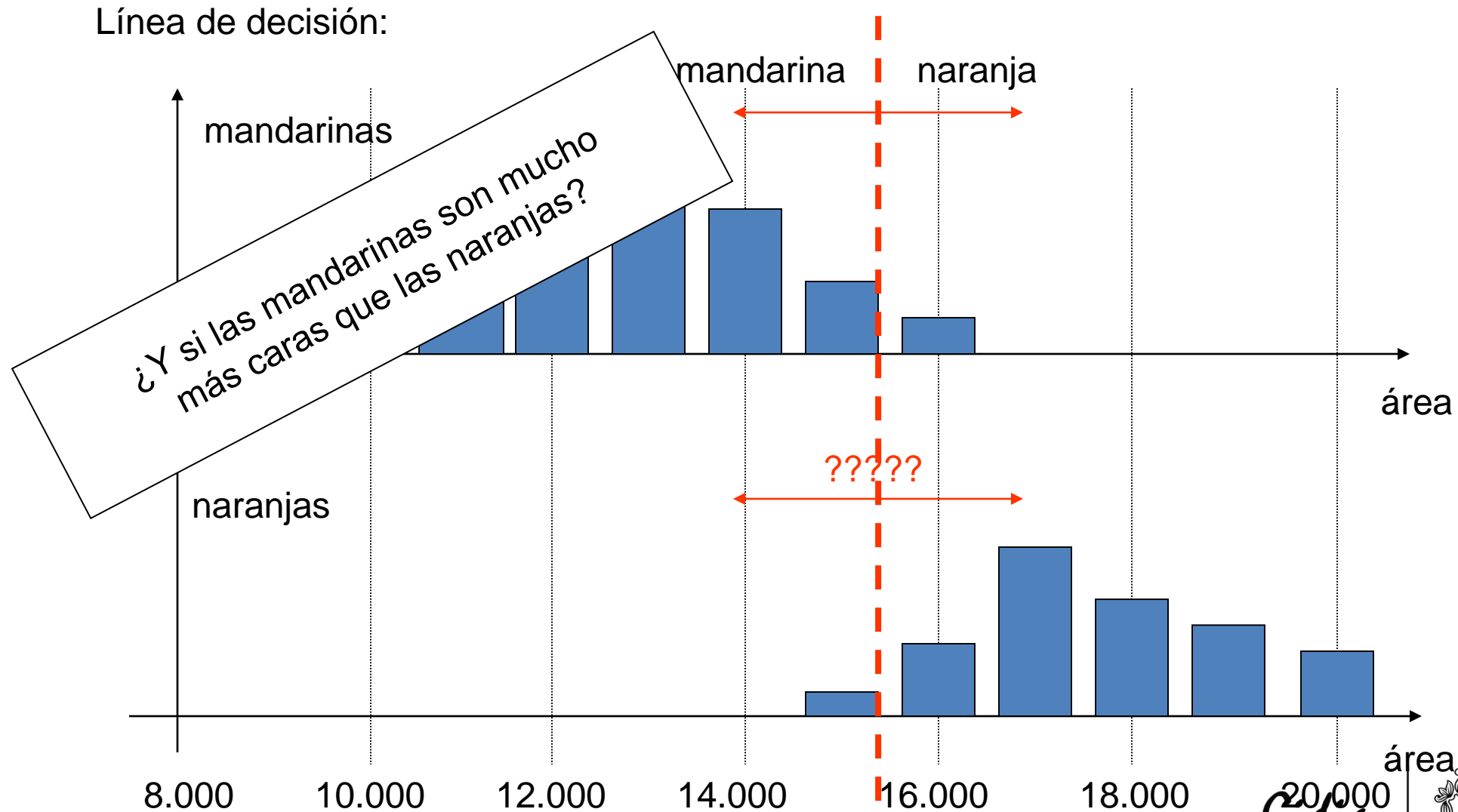
Línea de decisión:



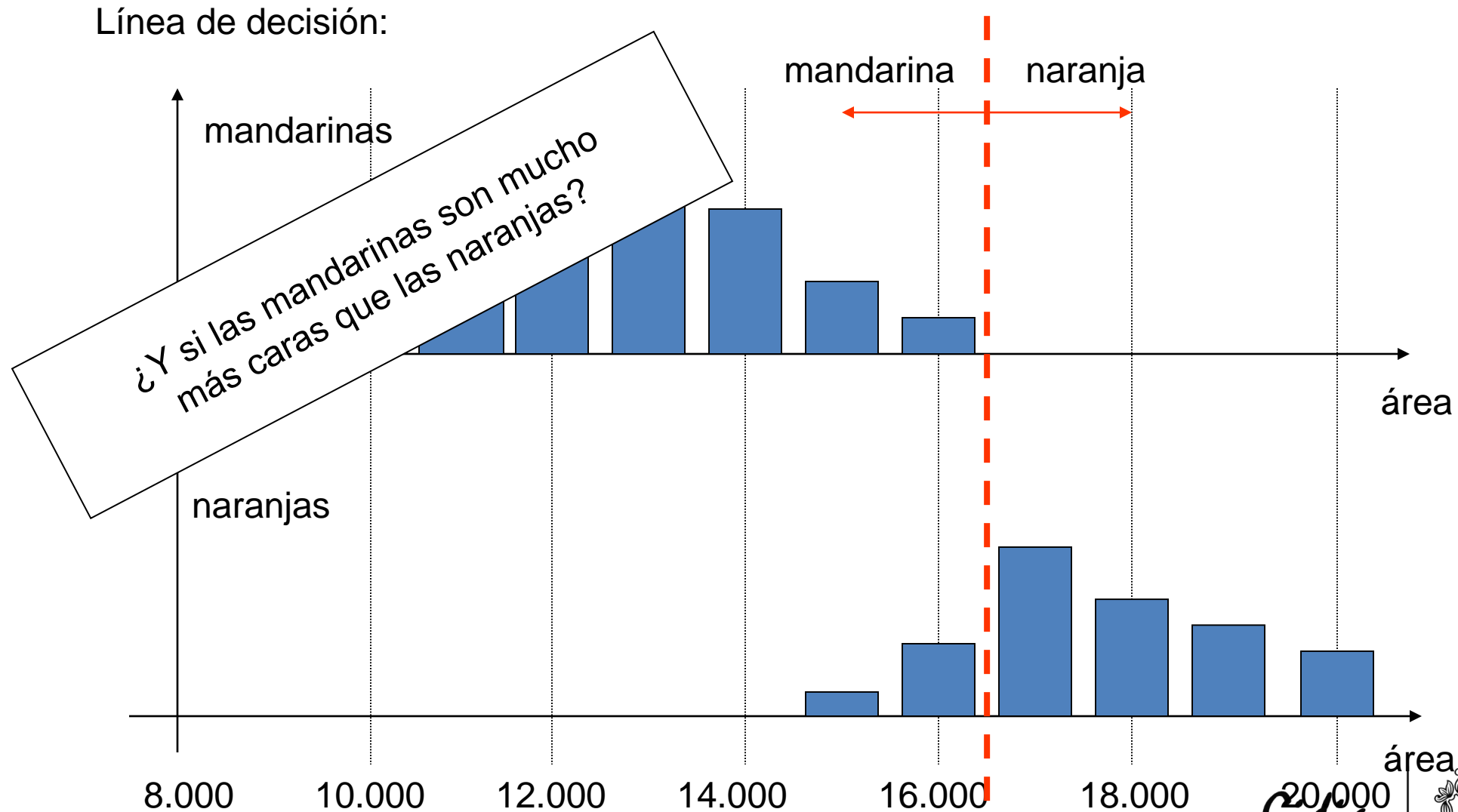


La separación es buena pero no es perfecta

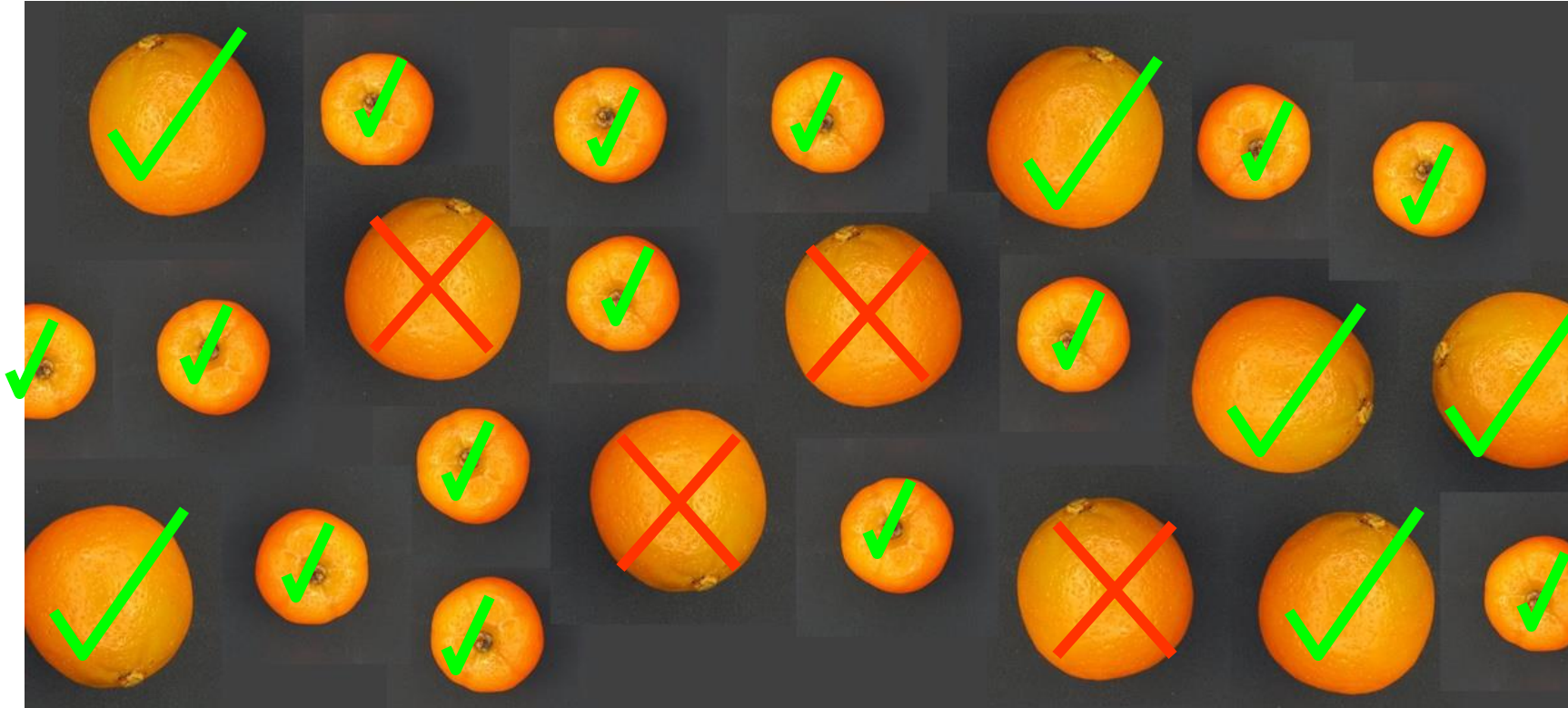
## Un ejemplo práctico



## Un ejemplo práctico



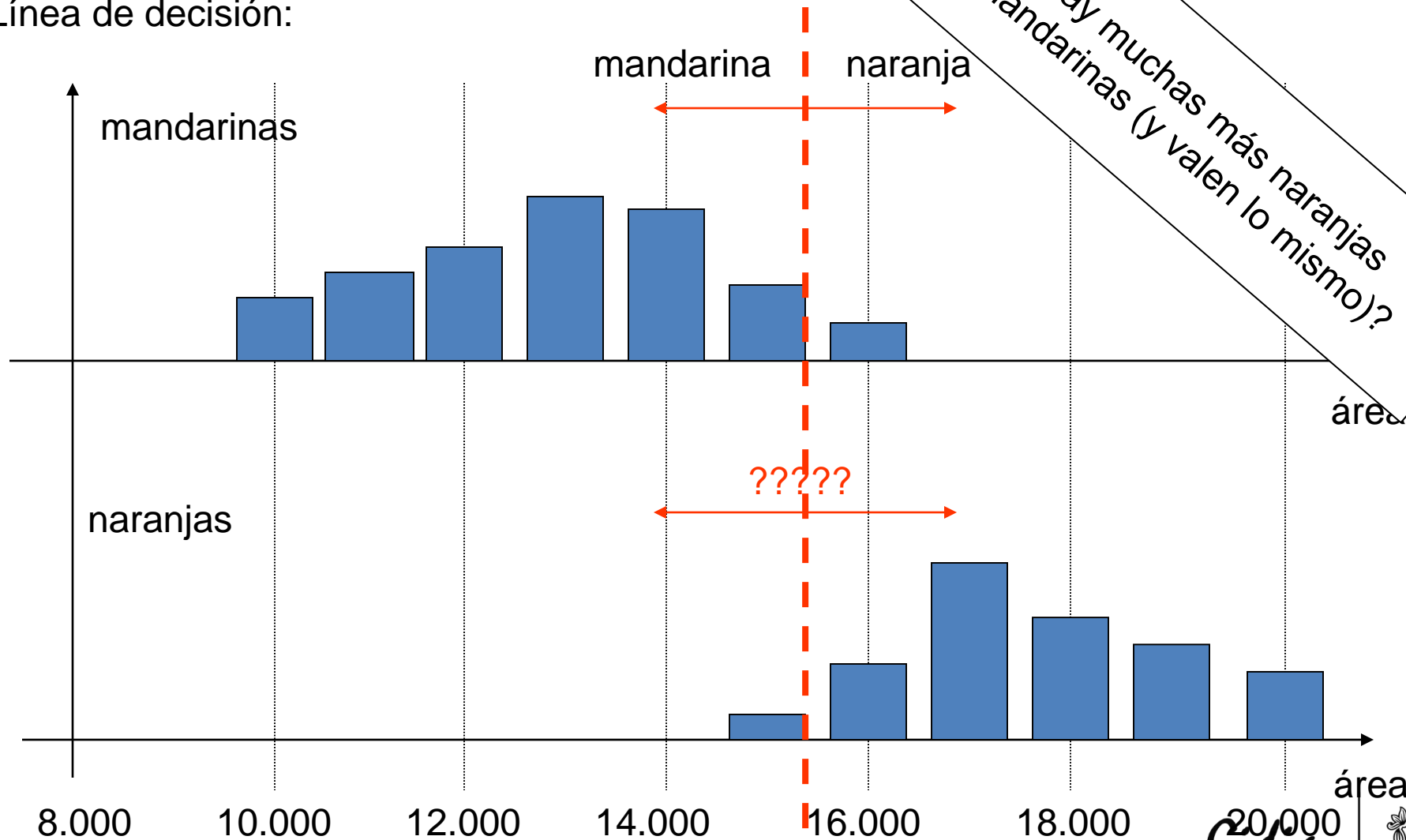
## Un ejemplo práctico



Todas las mandarinas son clasificadas perfectamente  
... pero el costo es que hay varias naranjas mal clasificadas

## Un ejemplo práctico

Línea de decisión:

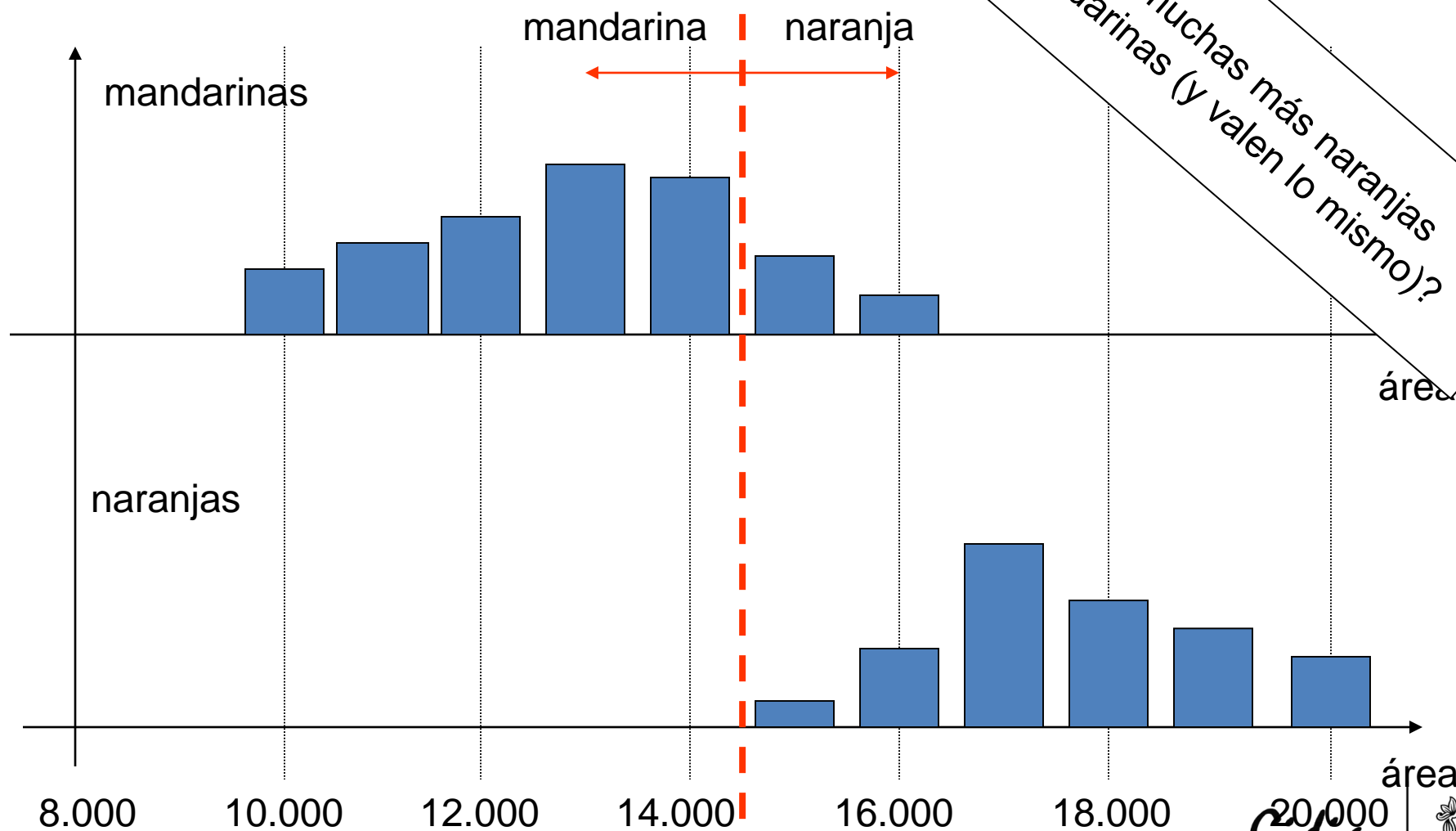


¿Y si hay muchas más naranjas que mandarinas (y valen lo mismo)?



## Un ejemplo práctico

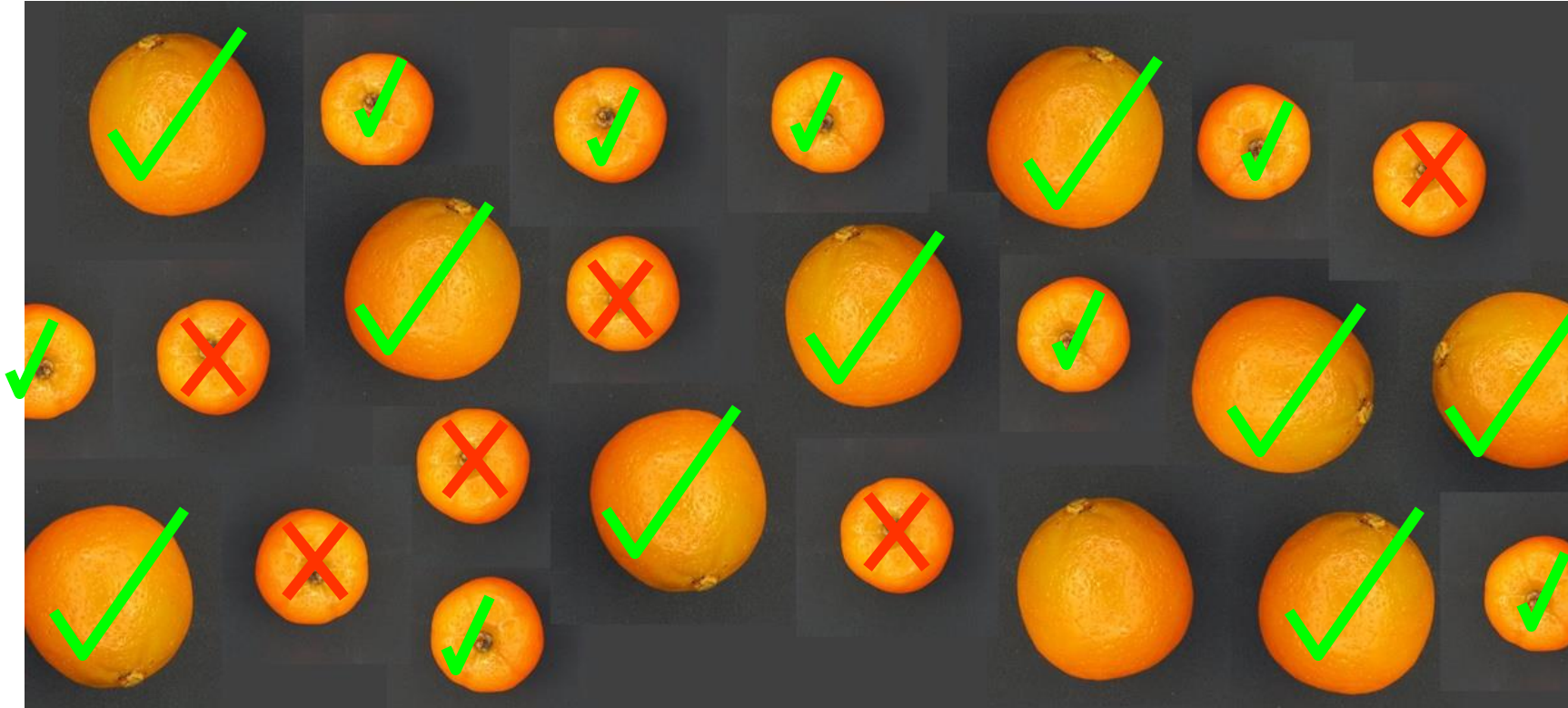
Línea de decisión:



¿Y si hay muchas más naranjas que mandarinas (y valen lo mismo)?

## Un ejemplo práctico

---



Todas las naranjas son clasificadas perfectamente  
... pero el costo es que hay varias mandarinas mal clasificadas

## Un ejemplo práctico

---

¿Cómo mejorar el desempeño?

Medición del color es una segunda alternativa: (*las naranjas son más verdes*)



Verde = 23.6%



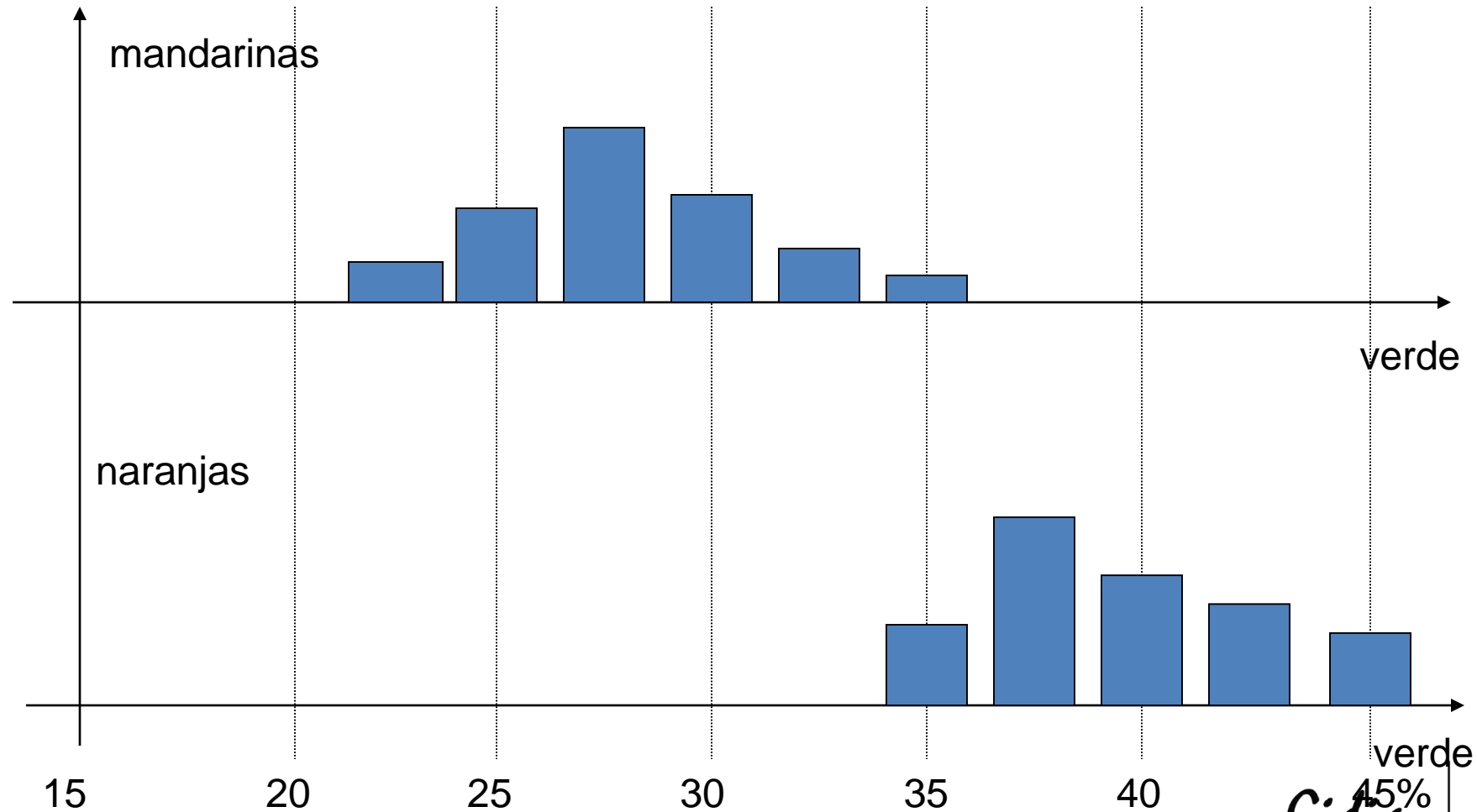
Verde = 46%

Recolección de información: Porcentaje de Verde

Naranja-01	41.3	Mandarina-01	23.6
Naranja-02	39.8	Mandarina-02	30.1
Naranja-03	36.5	Mandarina-03	37.1
Naranja-04	44.6	Mandarina-04	17.9
Naranja-05	41.2	Mandarina-05	19.7
Naranja-06	44.9	Mandarina-06	30.5
Naranja-07	44.4	Mandarina-07	35.4
:		:	
Naranja-75	38.7	Mandarina-50	33.6

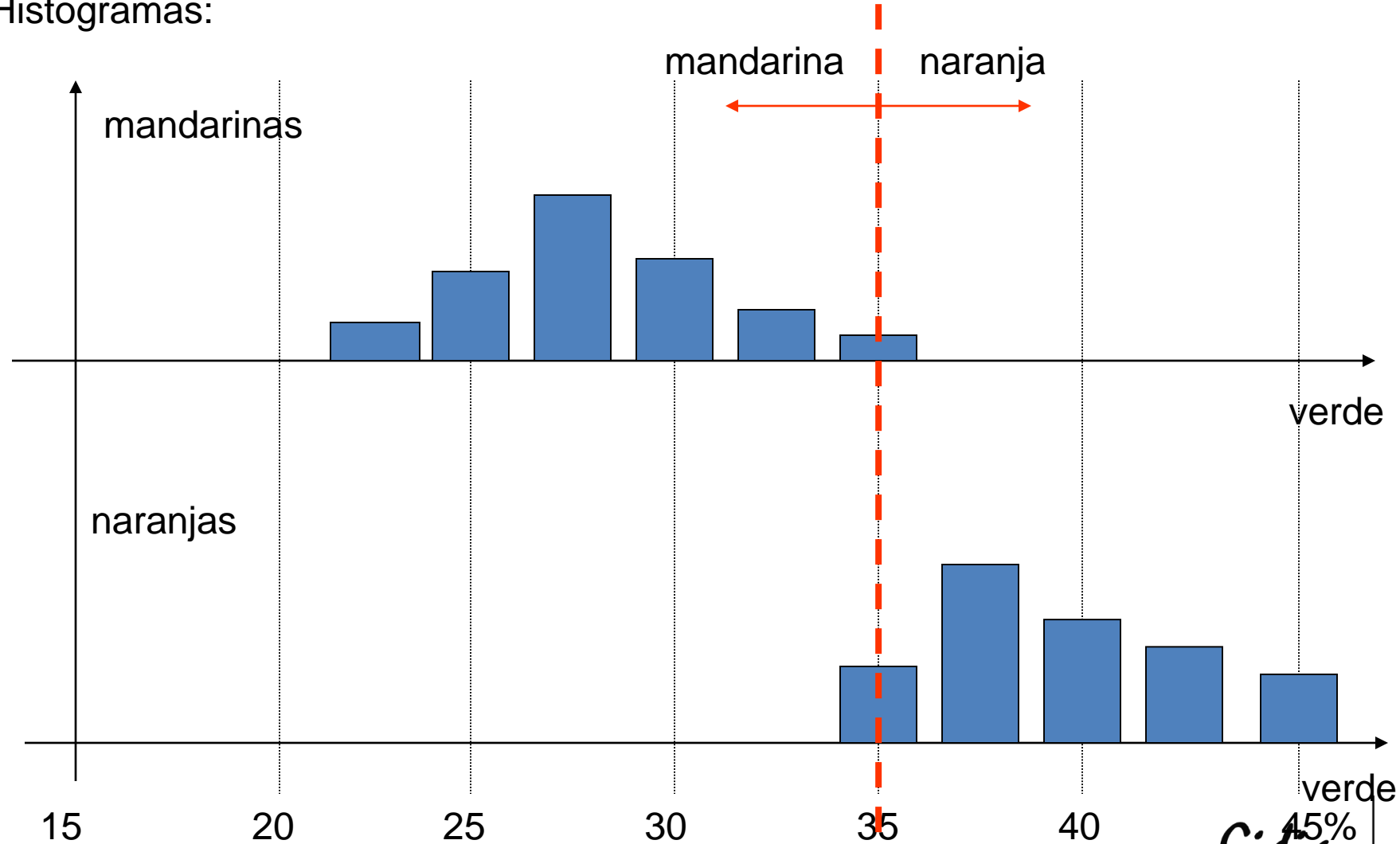
# Un ejemplo práctico

Histogramas:



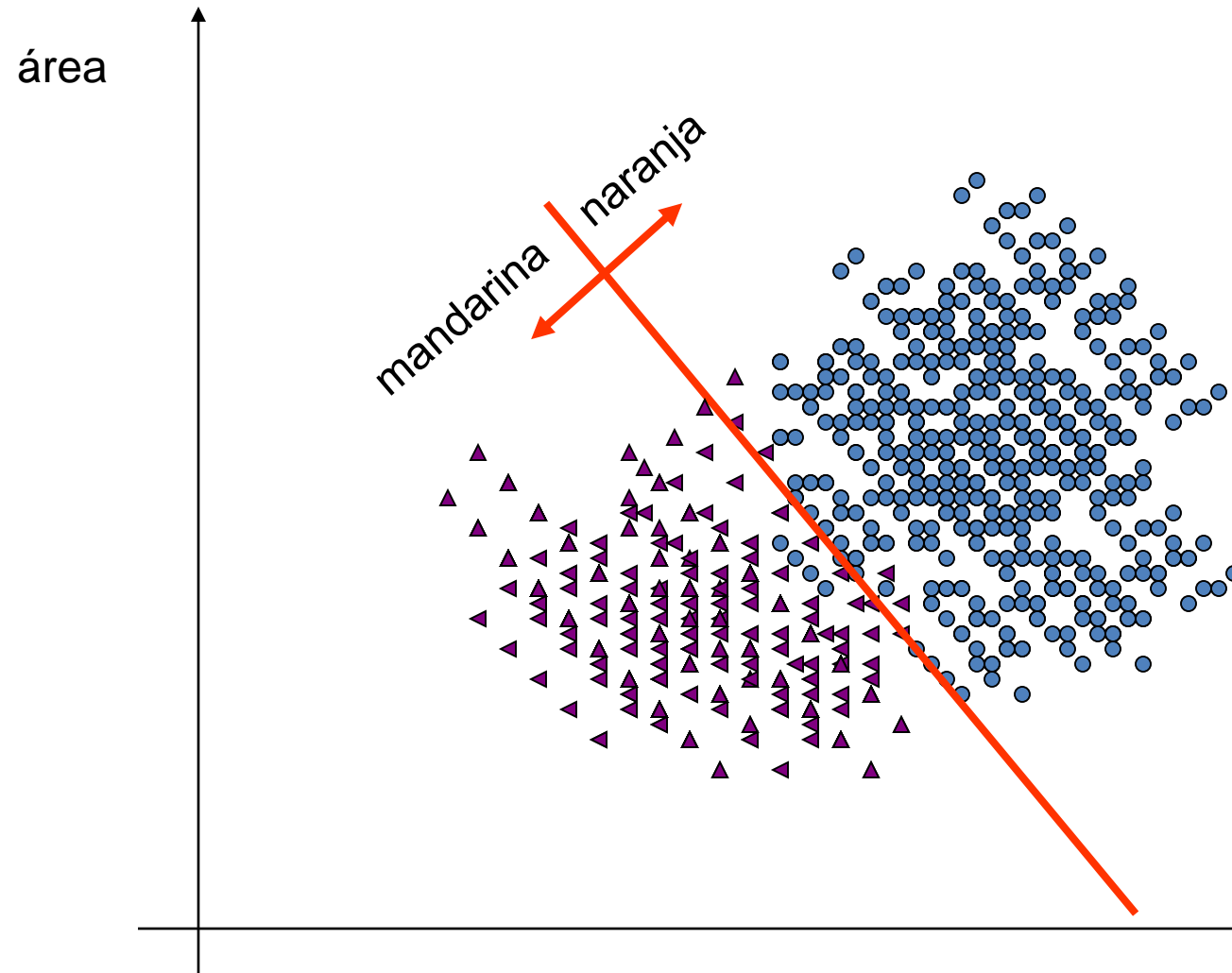
# Un ejemplo práctico

Histogramas:



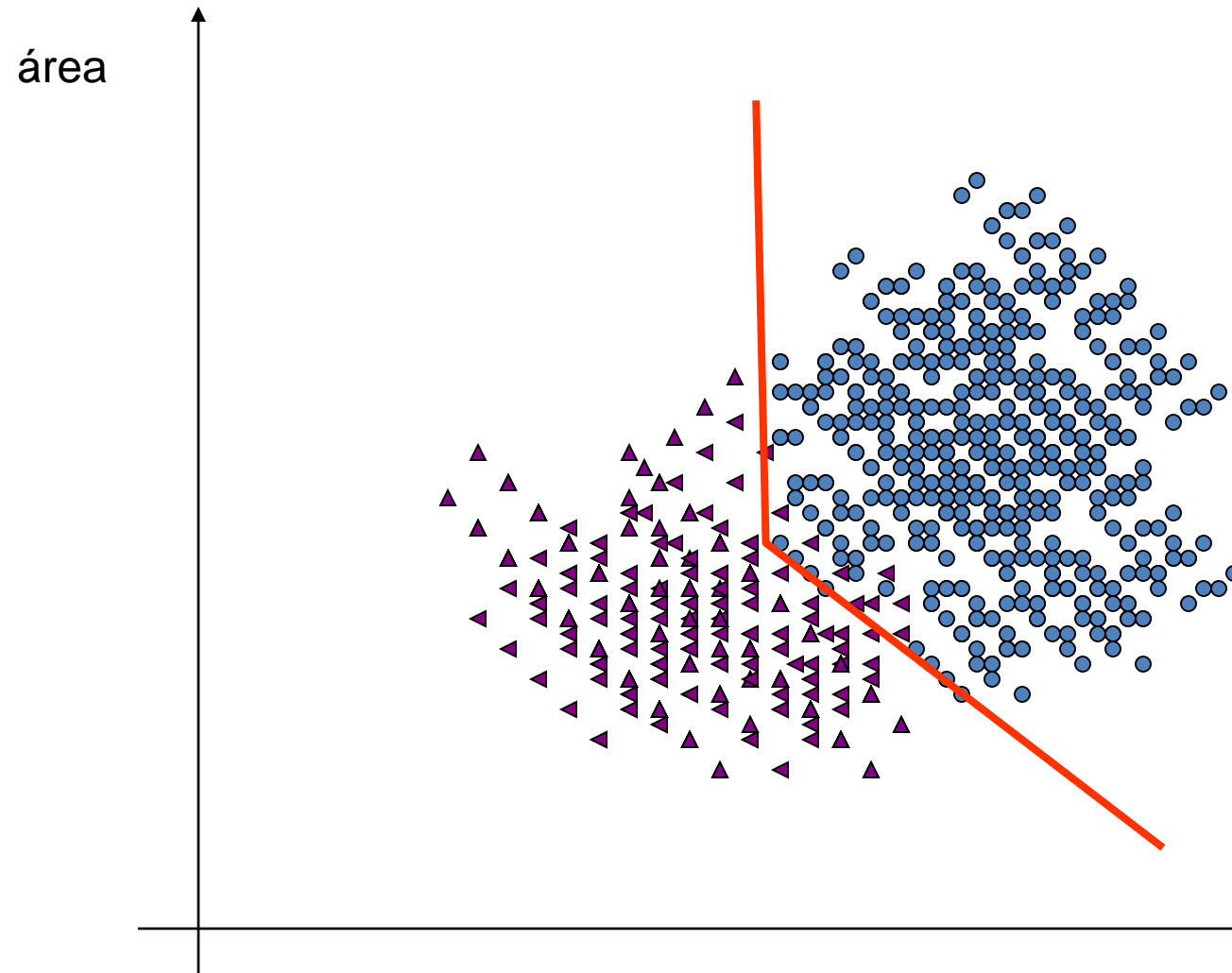
## Un ejemplo práctico

Uso de dos características:



## Un ejemplo práctico

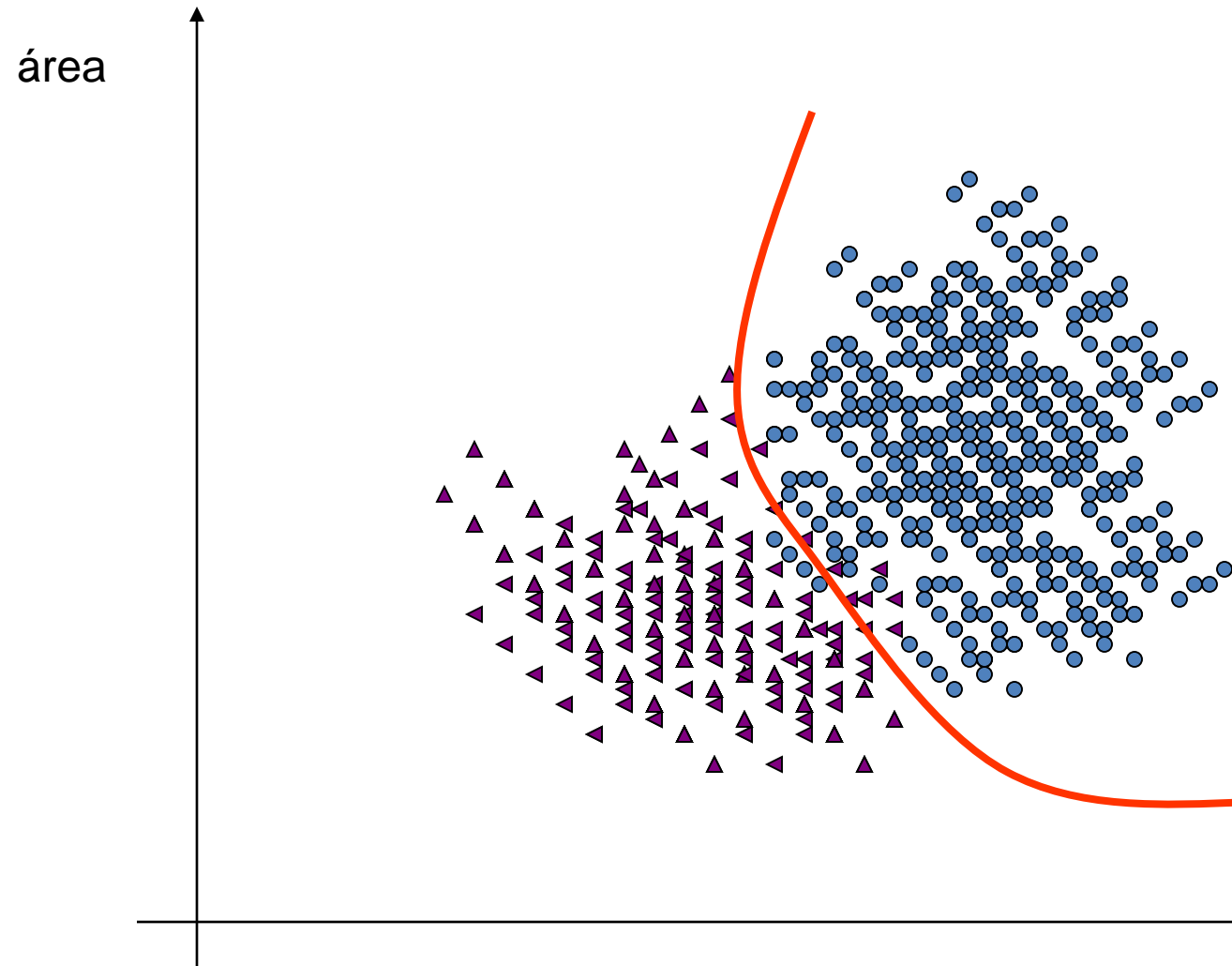
Uso de dos características:





## Un ejemplo práctico

Uso de dos características:



## Un ejemplo práctico

Uso de dos características:

