



UNIVERSIDAD
NACIONAL
DE COLOMBIA

APRENDIZAJE DE MÁQUINAS

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

<https://github.com/jwbranch/AprendizajeDeMaquinas>

<https://www.coursera.org/programs/unal-iuukt>

METODOLOGÍA ENSEÑANZA – APRENDIZAJE

El aprendizaje sincrónico involucra estudios online a través de una plataforma. Este tipo de aprendizaje sólo ocurre en línea. Al estar en línea, el estudiante se mantiene en contacto con el docente y con sus compañeros. Se llama aprendizaje sincrónico porque la plataforma permite que los estudiantes pregunten al docente o compañeros de manera instantánea a través de herramientas como el chat o el video chat.

Sesiones Remotas vía Google.Meet Sincrónicas y Asincrónicas

El aprendizaje asincrónico puede ser llevado a cabo online u offline. El aprendizaje asincrónico implica un trabajo de curso proporcionado a través de la plataforma o el correo electrónico para que el estudiante desarrolle, de acuerdo a las orientaciones del docente, de forma independiente. Un beneficio que tiene el aprendizaje asincrónico es que el estudiante puede ir a su propio ritmo.

Descripción del Curso

El curso introduce los conceptos fundamentales y los métodos más utilizados en el campo del aprendizaje de máquinas enfocados desde las perspectivas de la naturaleza del problema que se requiere resolver, esto es, aprendizaje supervisado orientado a los problemas de clasificación y regresión para aplicaciones de predicción o pronóstico. Aprendizaje no supervisado orientado a tareas de agrupar o etiquetar un conjunto de datos, También se incluyen la aproximación general de técnicas modernas de aprendizaje tales como el aprendizaje por refuerzo y aprendizaje profundo.

Contenido

1. **Introducción.**
2. Los datos en Aprendizaje de Máquinas.
3. Aprendizaje Supervisado.
4. Aprendizaje NO Supervisado.
5. Aprendizaje por Refuerzo.
6. Aprendizaje con Clases Desbalanceadas y Combinación de Modelos.
7. Aplicaciones y Casos de Éxito.

Bibliografía Recomendada

Oswaldo Simeone (2018), “A Brief Introduction to Machine Learning for Engineers”, Foundations and TrendsR in Signal Processing: Vol. 12, No. 3-4, pp 200–431. DOI: 10.1561/20000000102.

Goodfellow, I., Bengio, Y. y Courville, A. (2016) Deep Learning, MIT Press.

Murphy, K. (2012). Machine Learning: A Probabilistic Perspective, MIT Press .

Hastie, T., Tibshirani, R. y Friedman, J. (2011). The Elements of Statistical Learning. Springer.
(Available for download on the authors' web-page.)

Szepesvári, C. (2010). Algorithms for Reinforcement Learning. Morgan and Claypool.

Haykin, S. (2008). Neural Networks and Learning Machines. Pearson.

Sutton, R. y Barto, A. (1998). Reinforcement Learning: An Introduction. MIT Press.

Mitchell, T. M. (1997). Machine Learning. 1st. McGraw-Hill Higher Education. (Chapter 1)

EVALUACIÓN



Certificado Coursera

Sesenta años de inteligencia artificial – UNAM (Obligatorio)

IA para todos – Andrew Ng (Obligatorio)

Structuring Machine Learning Projects – Andrew Ng (Obligatorio)

20%

(Máx. 31 de Dic/2020)

Machine Learning - University of Washington → **Curso #1: Machine Learning Foundations: A Case Study Approach** (Obligatorio)

Informe de Lectura (Individual)

Machine Learning Algorithms: A Review

Machine Learning aspects and its Applications Towards Different Research Areas

20%

(Máx. 21 de Nov/2020)

Trabajo Final (Debe ser en Grupo 3 ó 5 personas)

Obtener el conjunto de datos (texto o audio o video o imagen) de los siguientes repositorios o cualquier otro disponible:

<http://www.ics.uci.edu/~mllearn/databases/>

<https://www.kaggle.com/datasets>

Origen, atributos, clases

"Scatter plot" de los datos

Visualización del conjunto en 2D (PCA o MDS)

Seleccionar un método de Entrenamiento y Evaluar el Desempeño.

60%

(Máx. 21 de Dic/2020)

APRENDIZAJE DE MÁQUINAS

Introducción

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

<https://github.com/jwbranch/AprendizajeDeMaquinas>

<https://www.coursera.org/programs/unal-iuukt>

Motivación

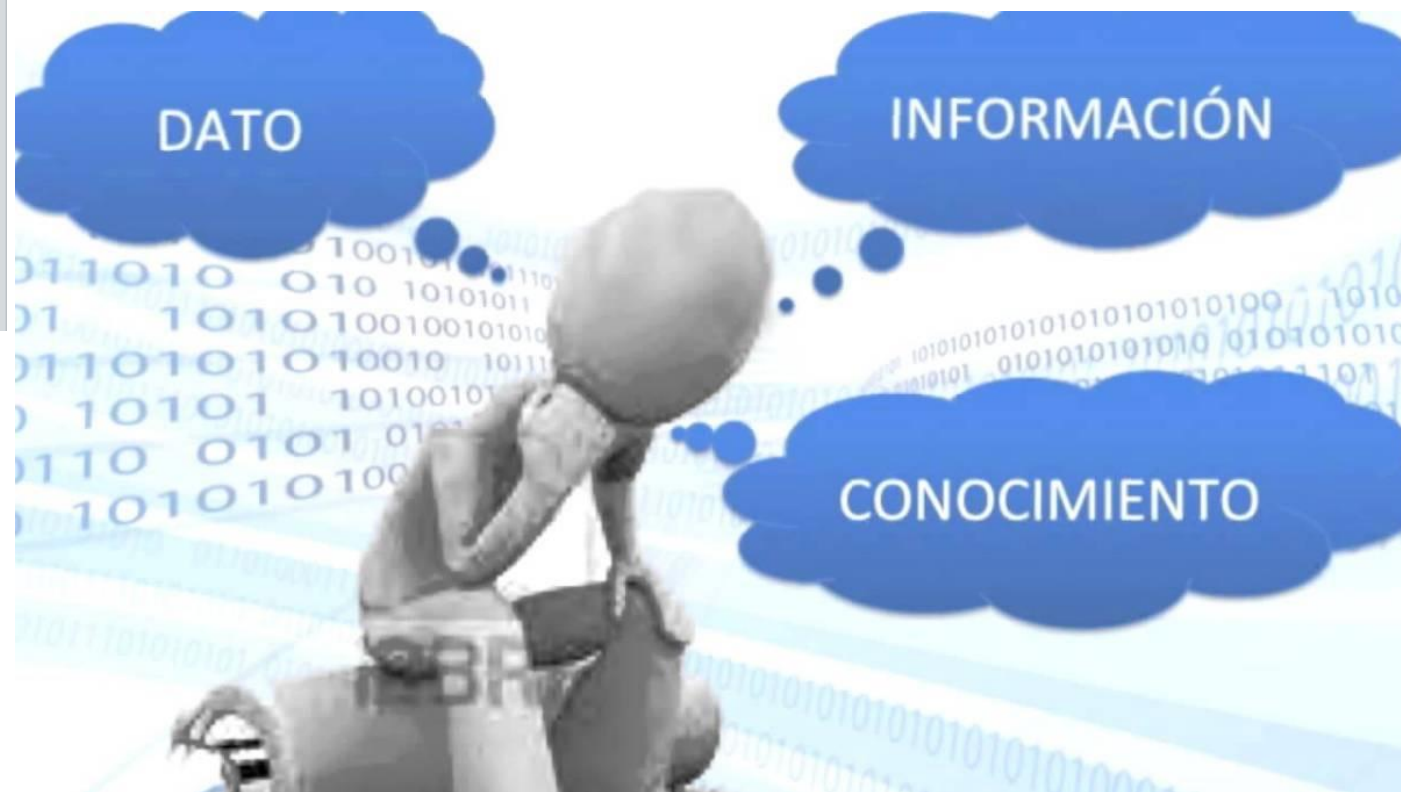
OBSERVE EL VIDEO Y RESPONDA A LAS SIGUIENTES PREGUNTAS:

-  ¿Cuántos datos se requieren para entrenar un sistema de visión artificial?
-  ¿Es posible decir que los computadores ya sobrepasaron la capacidad humana?
-  ¿Qué problemas evidencian los sistemas de visión artificial, y en general de los sistemas de Reconocimiento de Patrones?

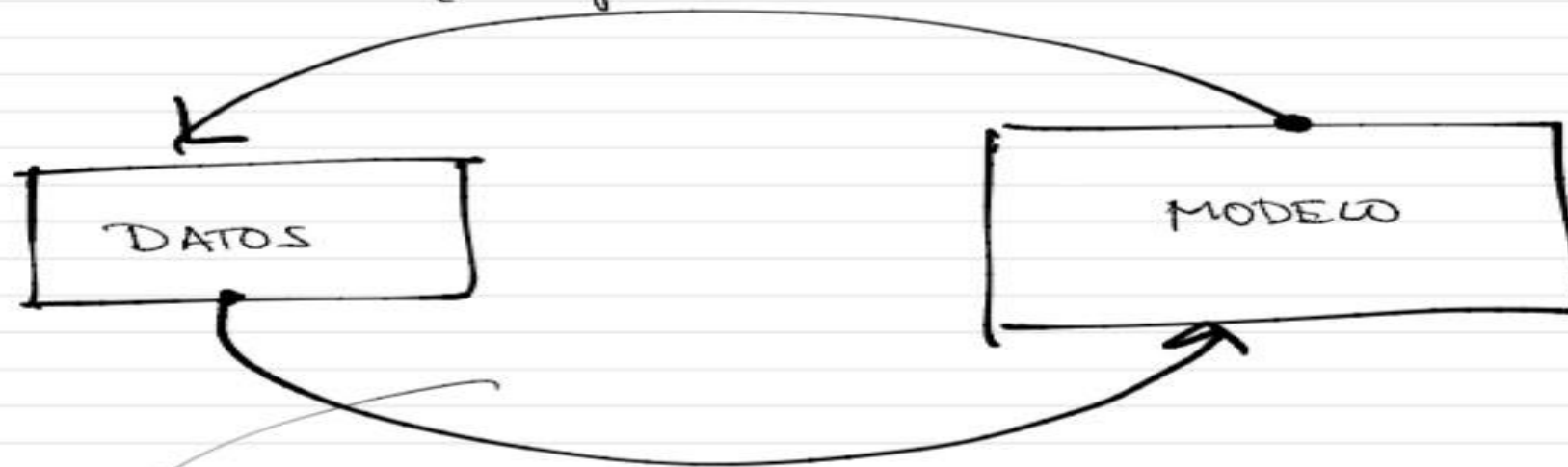


<https://www.ted.com/talks/fei-fei-li-how-we-re-teaching-computers-to-understand-pictures?language=es>

EL PROCESO



Aprox Estadística
Se postula el modelo
y luego se lleva al mundo real



Aproximación
IA basada en alguna
medida de la
precisión.

(Se corrige si el modelo no
es suficientemente preciso)

círculo virtuoso

- Los datos llevan al modelo
- Si el modelo no tiene la precisión se estudian sus propiedades para adaptarlo a como funciona el mundo real y tener más precisión

¿Qué es el aprendizaje?

El aprendizaje es el proceso de adquirir conocimiento, habilidades, actitudes o valores, a través del estudio, la experiencia o la enseñanza; dicho proceso origina un cambio persistente, medible y específico en el comportamiento de un individuo y, según algunas teorías, hace que el mismo formule un constructo mental nuevo o que revise uno previo. (*Wikipedia*)

”Find a bug in a program, and fix it, and the program will work today. Show the program how to find and fix a bug, and the program will work forever.”

Oliver G. Selfridge, AI's Greatest Trends and Controversies

Problemas del Aprendizaje

Programas o sistemas que aprenden: Sistemas capaces de optimizar un criterio de desempeño usando datos de ejemplo o experiencia pasada.

Se dice que un programa aprende de una experiencia E con respecto a una clase de tareas T y medida de desempeño P , si su desempeño en las tareas en T , de acuerdo con la medida P , mejora con la experiencia E .

Aprender a jugar damas

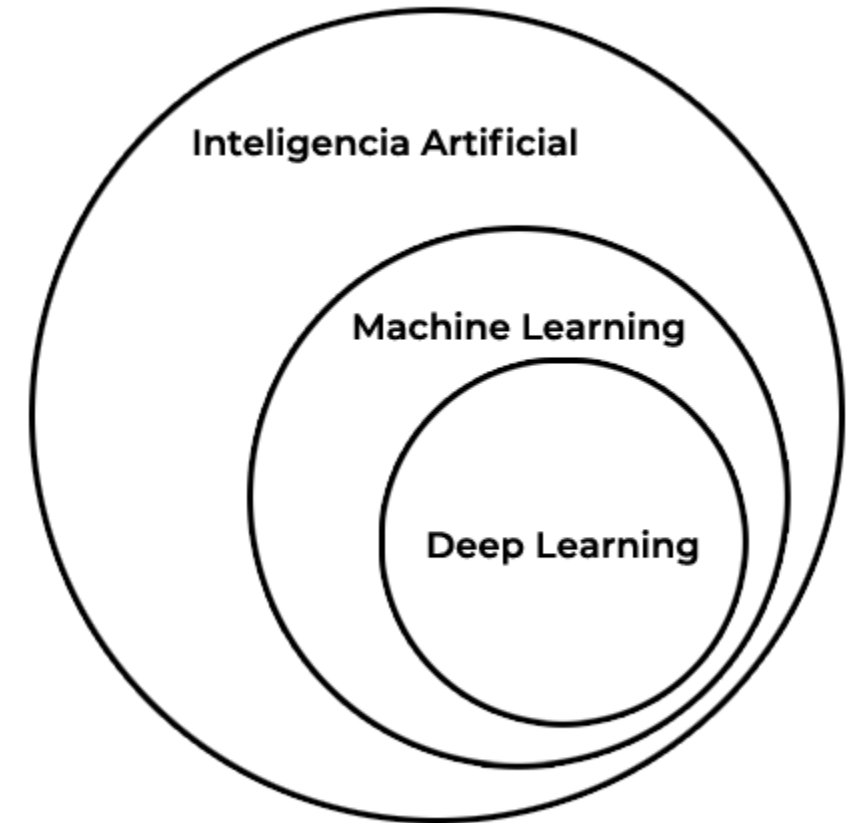
- **Tarea T :** jugar damas.
- **Medida de desempeño P :** porcentaje de juegos ganados contra oponentes.
- **Experiencia de entrenamiento E :** jugar partidas de práctica contra si mismo.

Aprender a reconocer escritura manuscrita

- **Tarea T :** reconocer y clasificar palabras manuscritas en imágenes.
- **Medida de desempeño P :** porcentaje de palabras correctamente clasificadas.
- **Experiencia de entrenamiento E :** una base de datos de palabras manuscritas clasificadas.

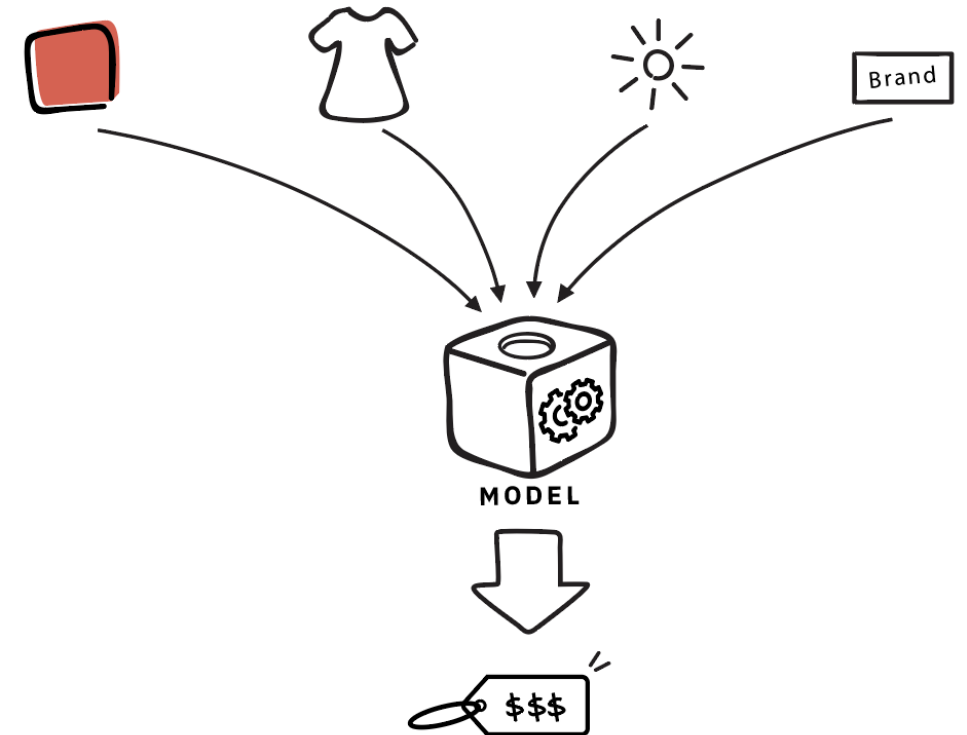
Aprendizaje de Máquinas

- “Es el subcampo de las ciencias de la computación y una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras *“aprendan”*.”
- “Es el desarrollo de sistemas que puedan cambiar su comportamiento de manera autónoma basados en su experiencia. El ML ofrece algunas de las técnicas más efectivas para el descubrimiento de conocimiento (patrones) en grandes volúmenes de datos.”
- A través de operaciones matemáticas y repetición de ejemplos, el algoritmo es capaz de aprender de la experiencia que está teniendo. Esto con el fin de mapear unos datos de entrada en una dimensión, a unos datos de salida en una dimensión diferente.

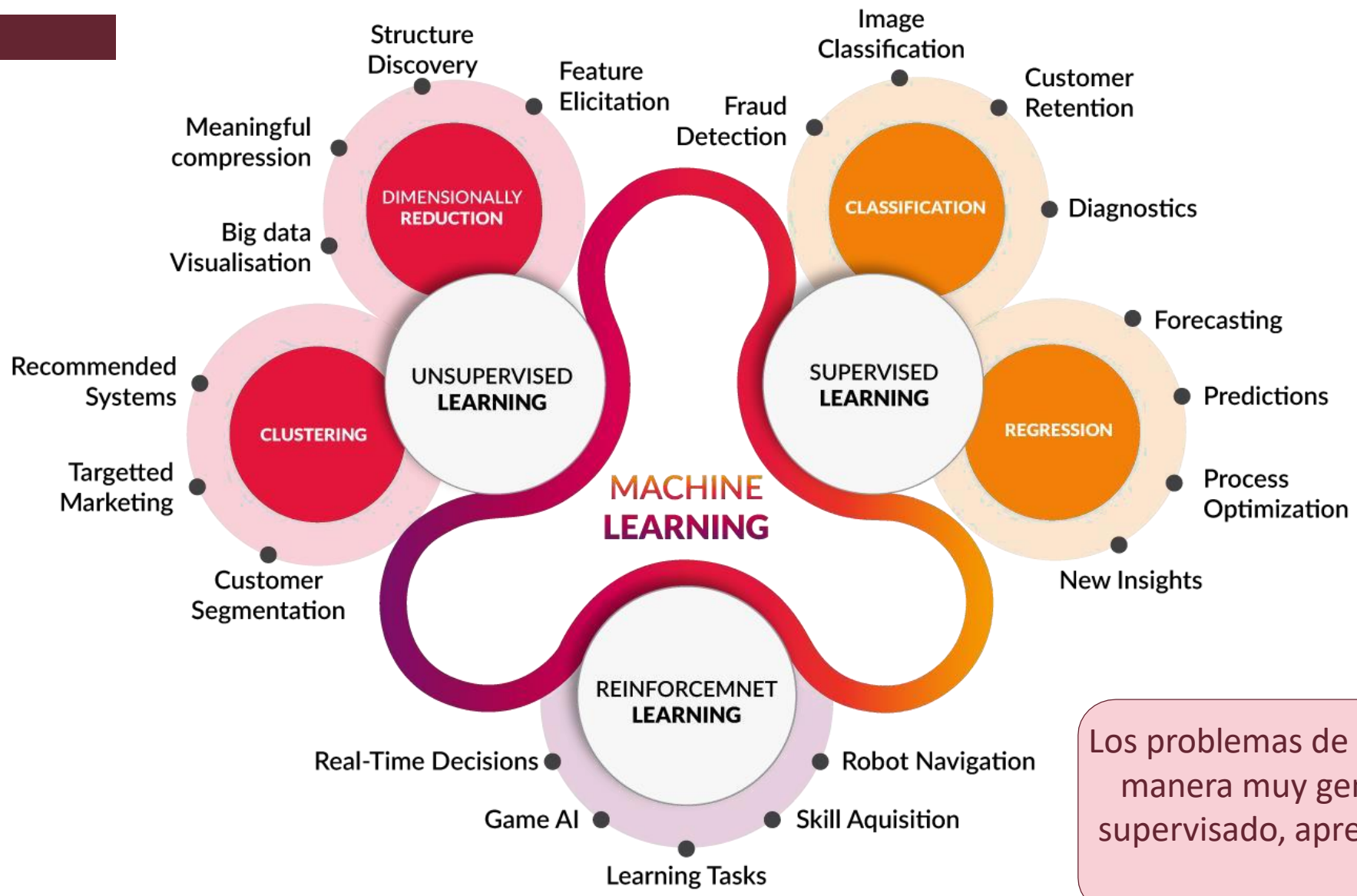


Aprendizaje de Máquinas y Predicción

- El problema central del aprendizaje de máquinas es la **predicción**, es decir, aplicar sobre datos nuevos un algoritmo que ha sido entrenado sobre un conjunto de datos históricos.
- Aunque suene como predecir el futuro, el término predicción generalmente se usa para el procesamiento de datos nuevos *in-situ*. Cuando los datos tienen un componente temporal se utiliza el término pronóstico.
- En este orden de ideas, cuando se habla de predicción se puede hacer referencia a:
 - **Clasificación** para obtener una etiqueta o clase conocida.
 - **Regresión** para obtener un valor numérico.
 - **Agrupamiento** para descubrir etiquetas o patrones nuevos a partir de datos no etiquetados a partir de medidas de asociación.



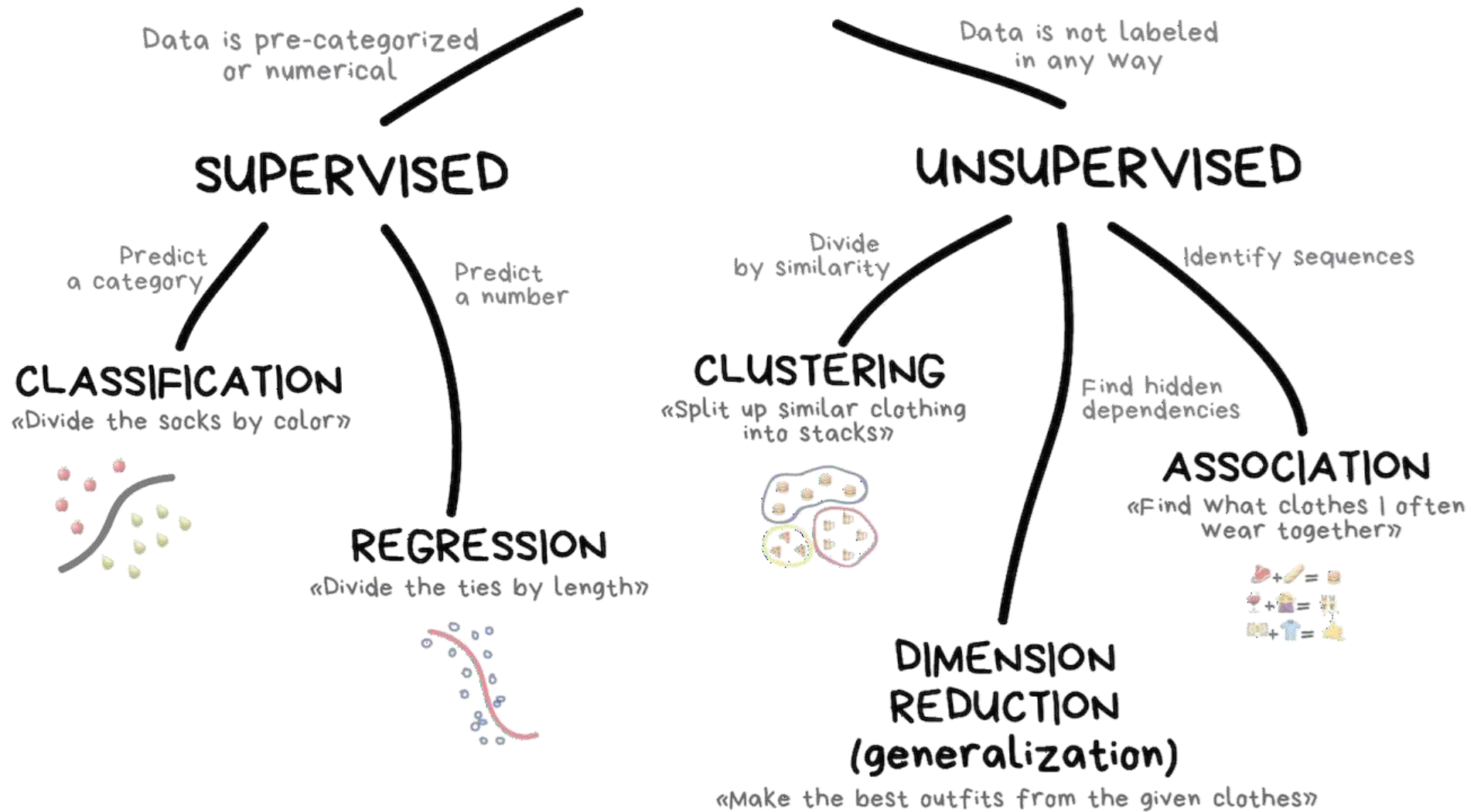
Tomada de: <https://medium.com/@srnghn>



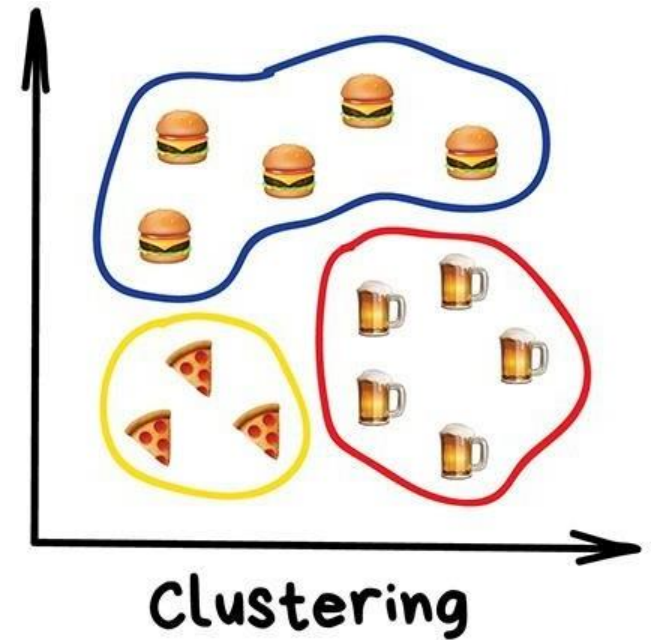
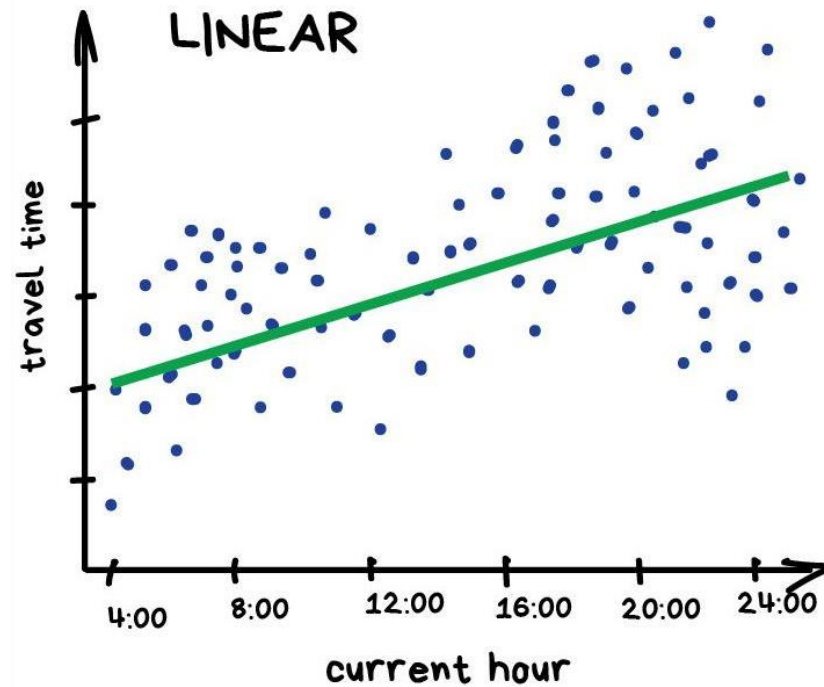
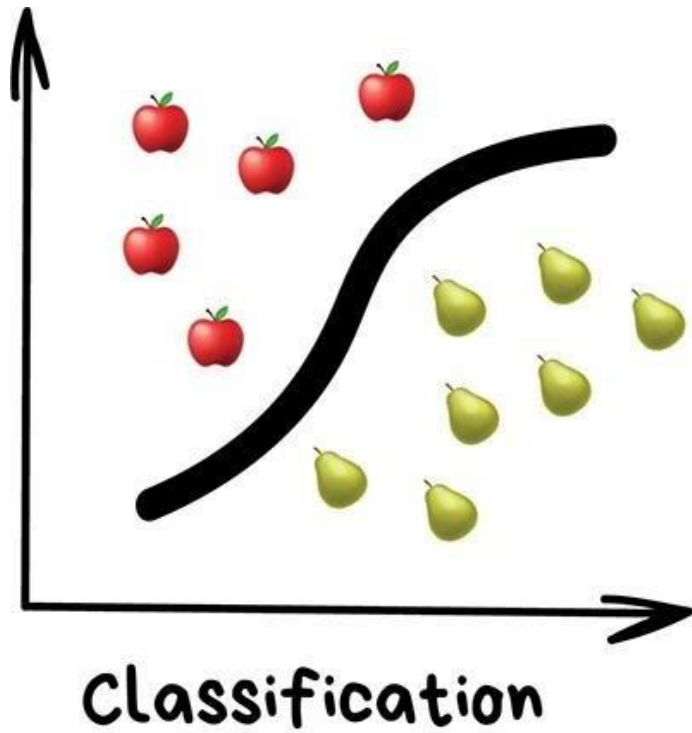
Los problemas de aprendizaje de máquina se dividen, de manera muy general, en tres categorías: aprendizaje supervisado, aprendizaje NO supervisado y aprendizaje por refuerzo.

Tomada de: <https://www.predictiveanalyticsworld.com/patimes/wp-content/uploads/2018/01/machinelearning-IMAGE.png>

CLASSICAL MACHINE LEARNING



Tipos de Aprendizaje



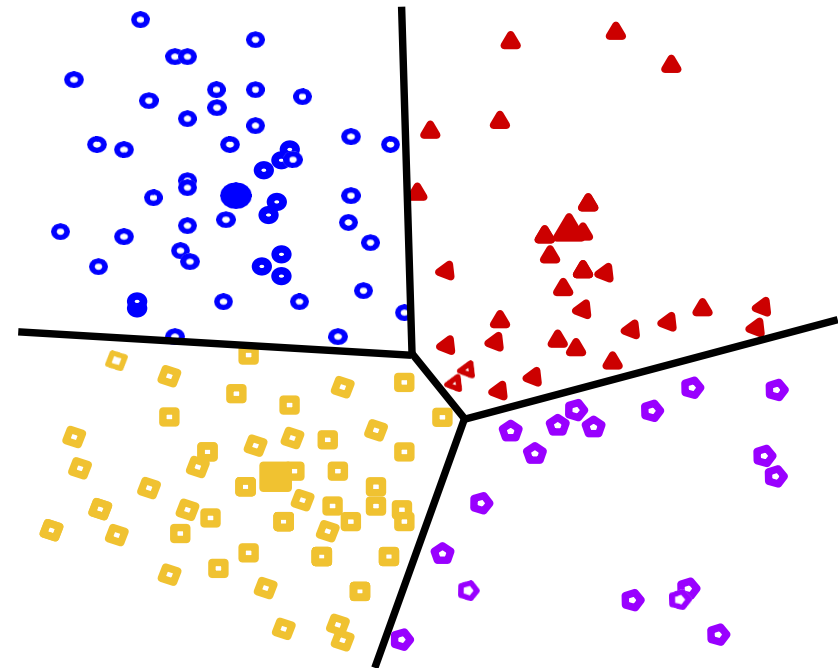
Aprendizaje Supervisado VS Aprendizaje NO Supervisado

Aprendizaje Supervisado: requiere de un conjunto de datos conocidos a partir del cual se crea un **modelo** para predecir el valor de una variable de salida. El aprendizaje supervisado se puede usar en dos tareas:

- **Clasificación:** en este caso la variable de salida es una etiqueta que determina la clase a la que pertenecen los datos de entrada, es decir, la variable de salida es una **variable discreta**.
- **Regresión:** en este caso los algoritmos de aprendizaje buscan predecir el valor de una **variable continua** a partir de los datos de entrada. Un ejemplo de una tarea de regresión es el de estimar la longitud de un salmón en función de su edad y su peso.

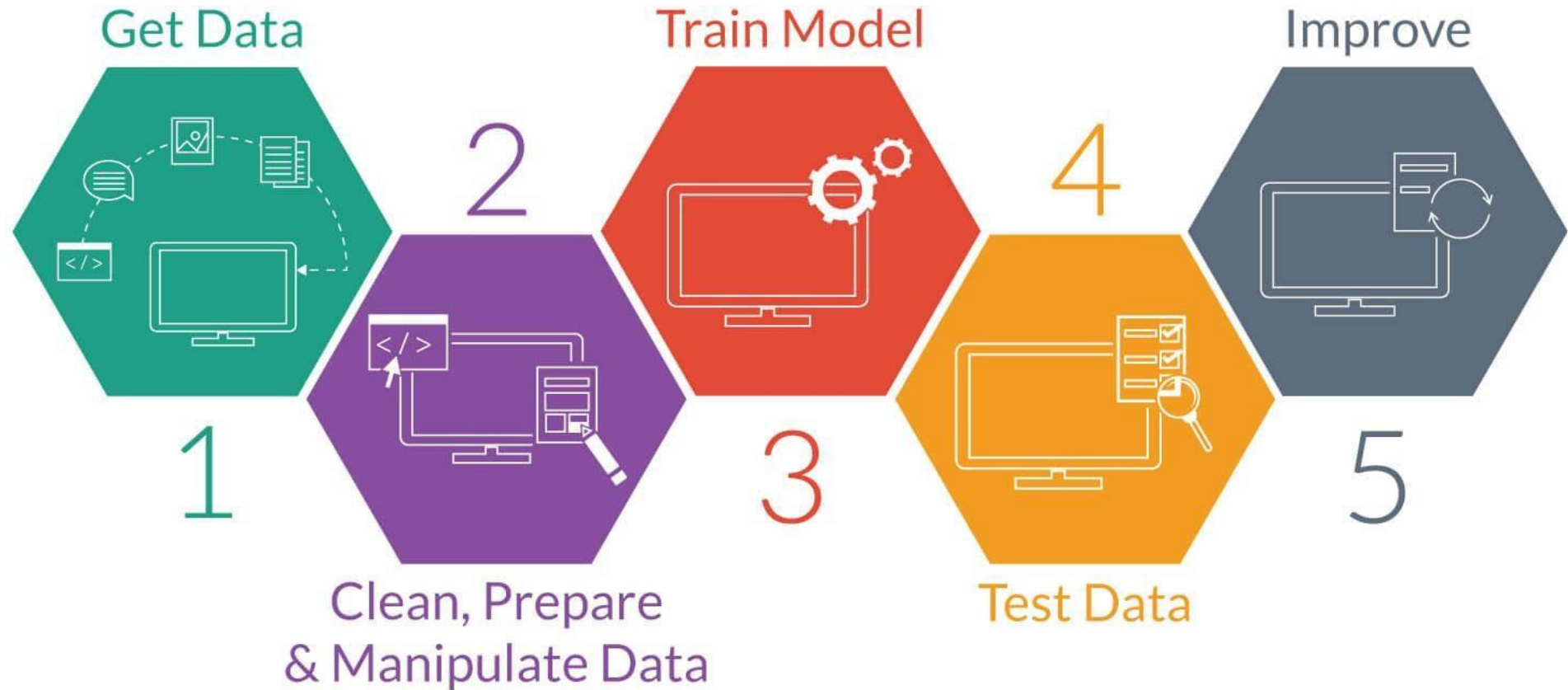
Aprendizaje Supervisado vs Aprendizaje NO Supervisado

Aprendizaje No Supervisado: se cuenta con un conjunto de datos de entrenamiento, pero no hay una variable específica de salida (se desconocen las clases). En este sentido, el objetivo de los problemas del aprendizaje no supervisado es, por ejemplo, el de agrupar los datos de entrada con base en algún criterio de similitud o disimilitud o determinar la distribución estadística de los datos, conocida como estimación de la densidad.



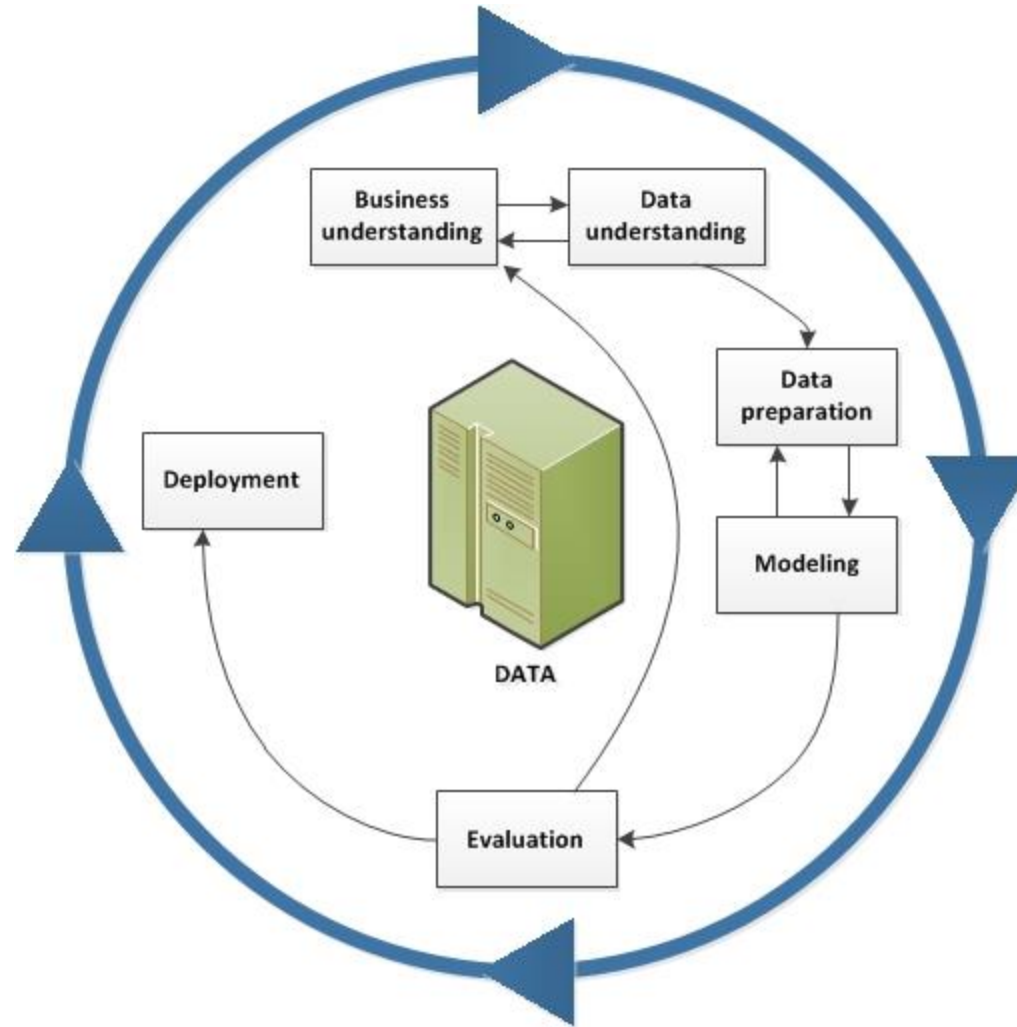
Metodología Clásica para el Desarrollo

<https://www.portalveterinaria.com/porcino/actualidad/31080/em-pig-data-em.html>



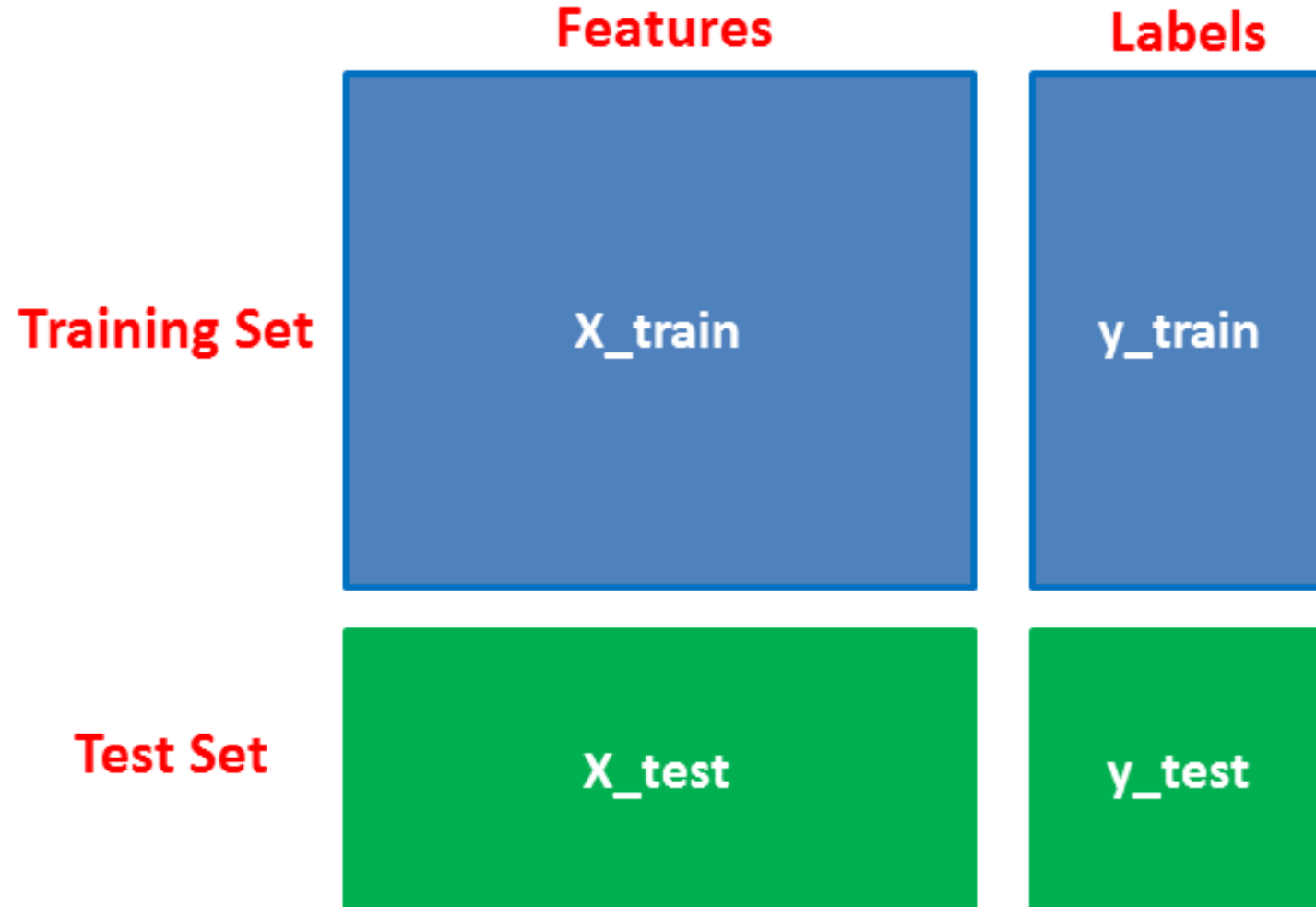
Tomada de: <https://newtiummedia.blob.core.windows.net/images/Steps-to-Predictive-Modelling.jpg>

Metodología iterativa: CRISP-DM



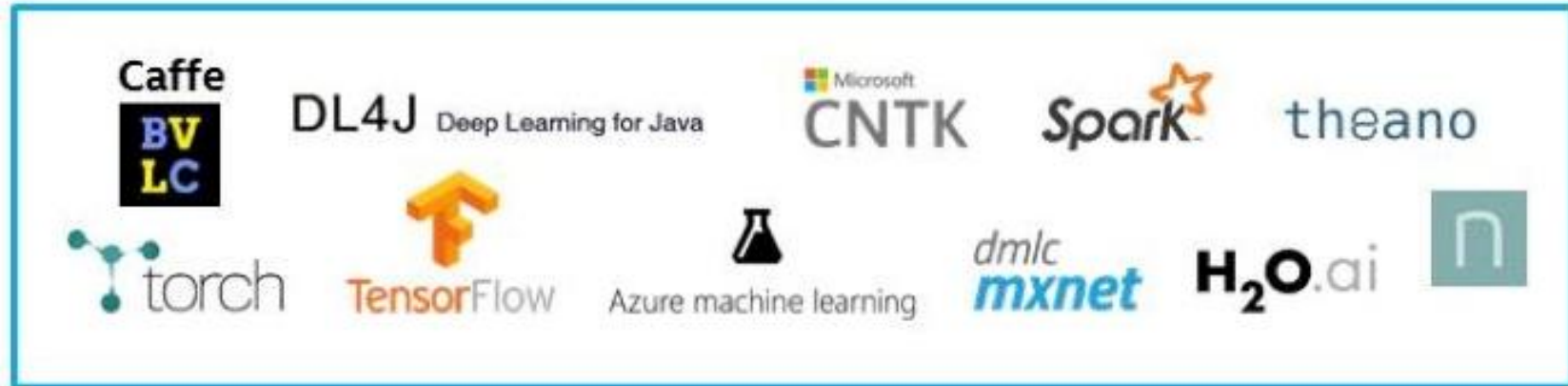
Tomada de: <https://www.ibm.com/>

Partición del Conjunto de Datos



Herramientas Tecnológicas

Top Frameworks



Programming languages



Tomada de: <https://datafloq.com/read/machine-learning-becomes-mainstream-advantage/2236>

Herramientas Tecnológicas



IP[y]: IPython
Interactive Computing



Machine Learning Overview

Machine Learning Basics

<https://youtu.be/ukzFI9rgwfU>

Machine Learning: Making Sense of a Messy World

<https://youtu.be/l95h4aIXfAA>

The 7 steps of machine learning

<https://youtu.be/nKW8Ndu7Mjw>

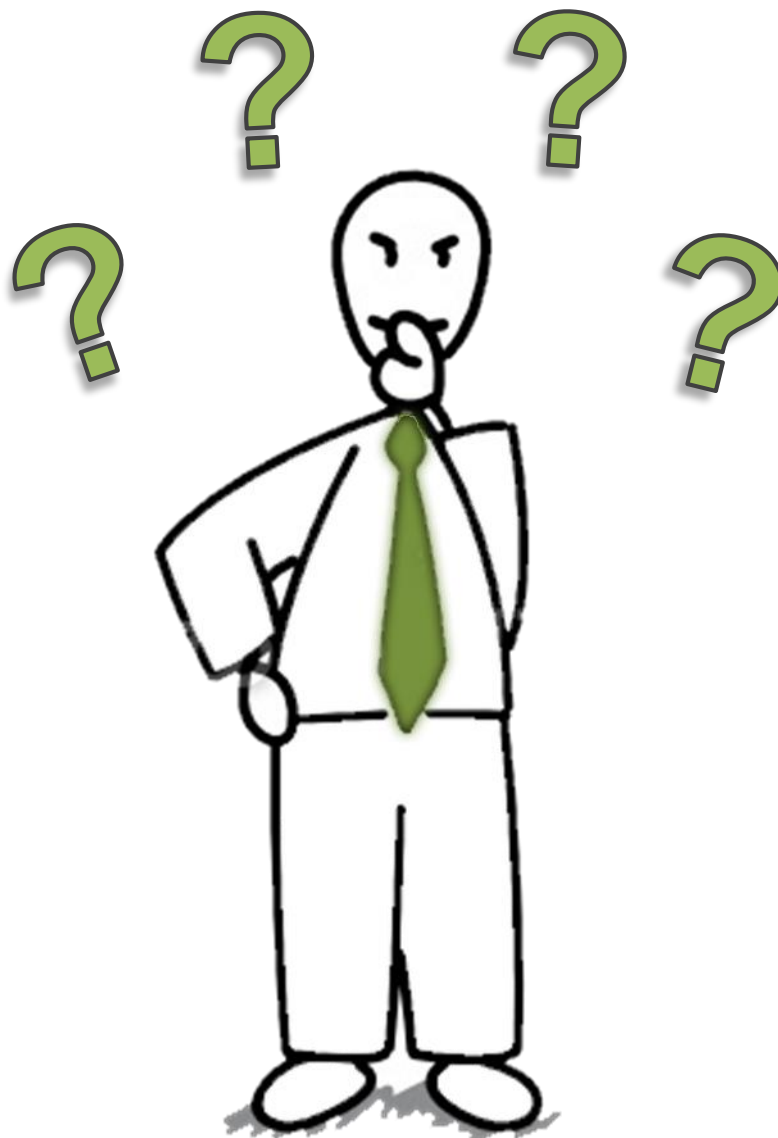
When Machine Learning Becomes AI

https://youtu.be/knazJ9_mRzE

Machine Learning Glossary

<https://developers.google.com/machine-learning/glossary/>

Preguntas





Reconocimiento de Patrones

Ejemplo Práctico - Mandarinas vs. Naranjas

[Capítulo 1]

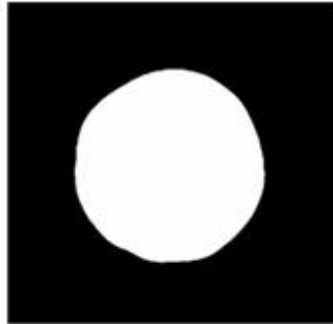
Domingo Mery

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

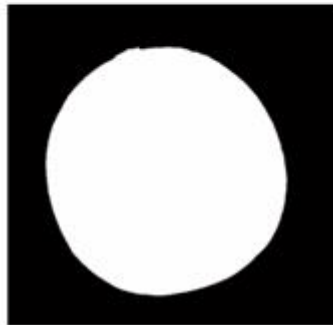


¿cómo separar las mandarinas de las naranjas?

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



Área = 15.457 pixeles

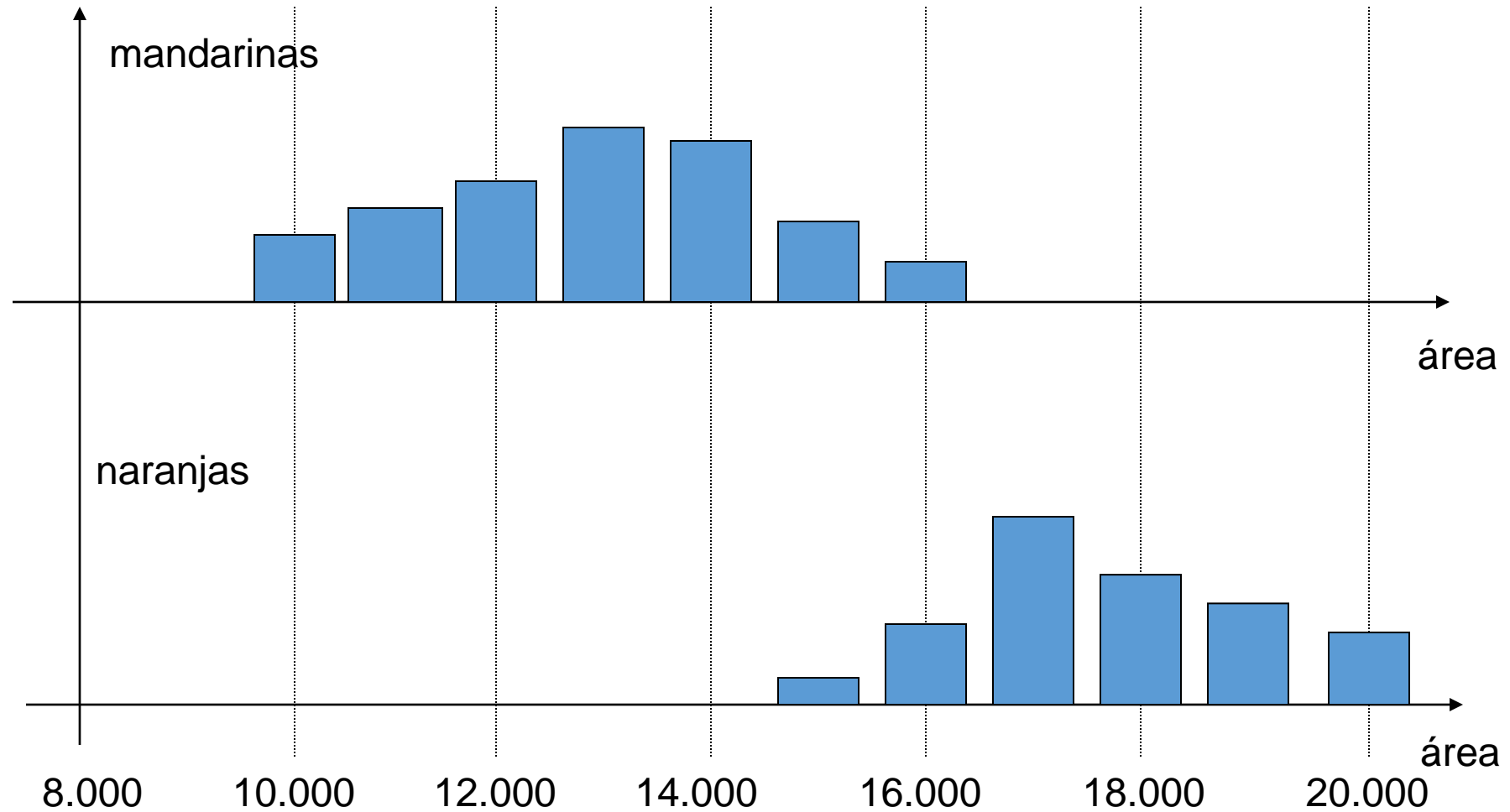


Área = 18.583 pixeles

Recolección de información: Área en Pixeles

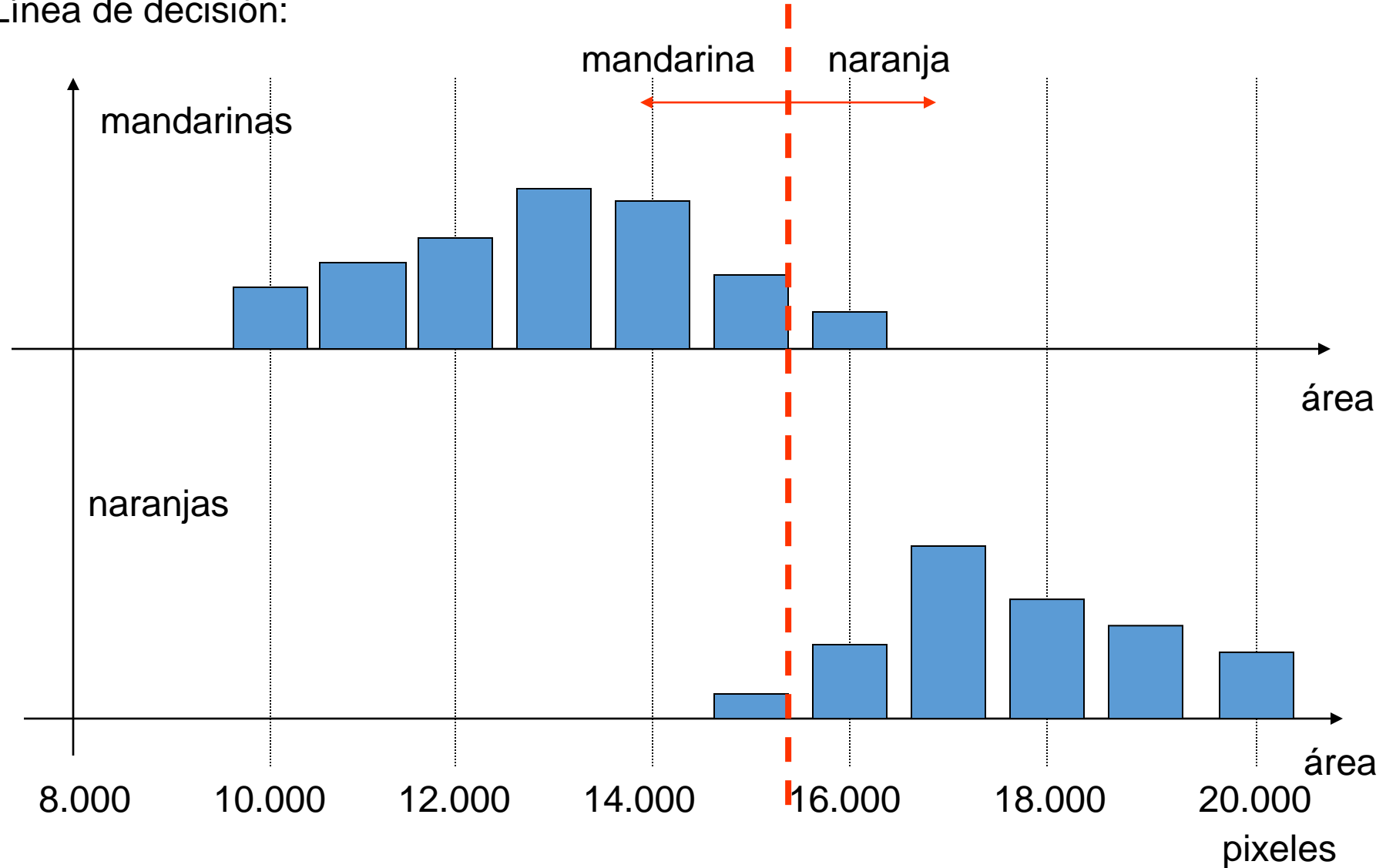
Naranja-01	19.327	Mandarina-01	13.221
Naranja-02	18.265	Mandarina-02	14.987
Naranja-03	17.456	Mandarina-03	15.321
Naranja-04	19.341	Mandarina-04	15.987
Naranja-05	16.342	Mandarina-05	16.345
Naranja-06	16.987	Mandarina-06	15.965
Naranja-07	17.001	Mandarina-07	16.341
:	19.056	:	
Naranja-75	15.900	Mandarina-50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

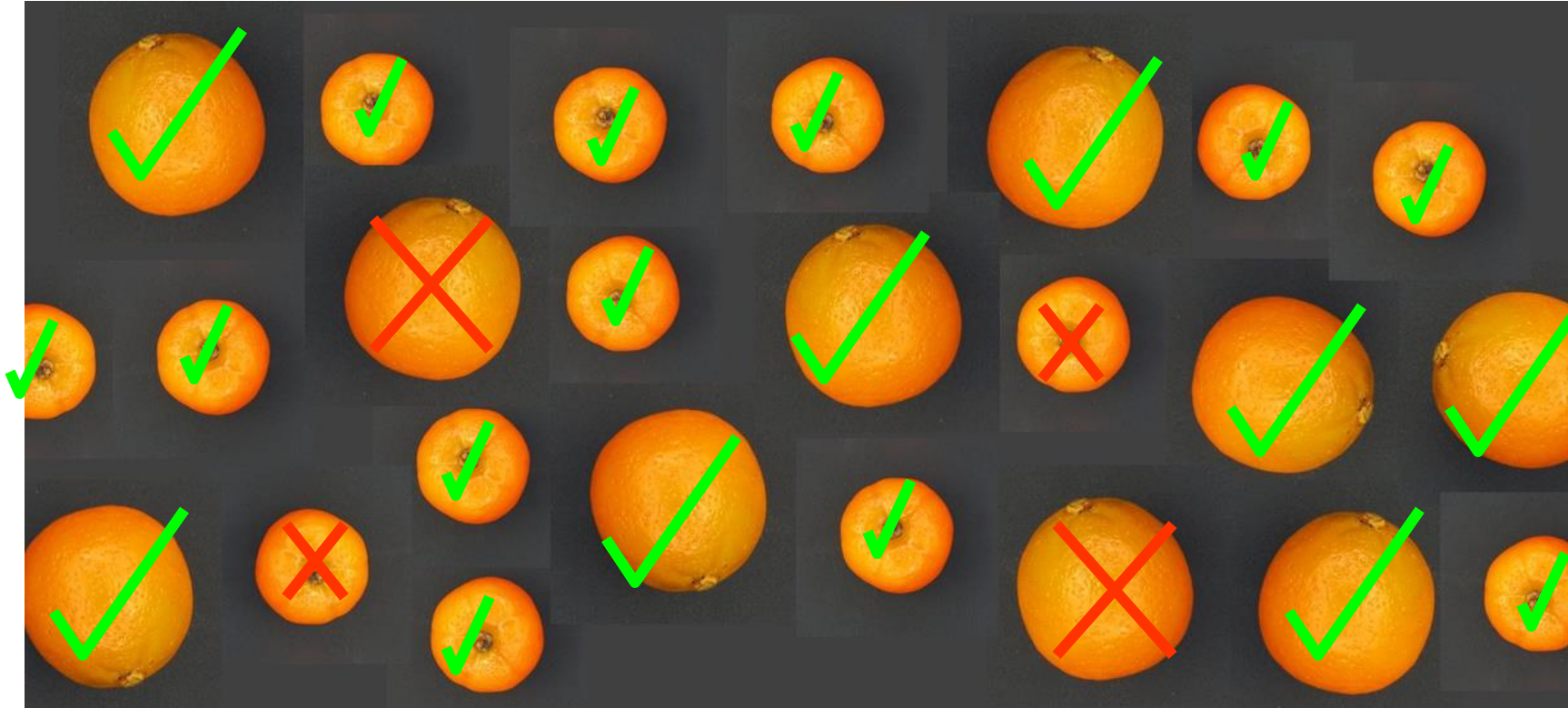
Histogramas:



Un ejemplo práctico

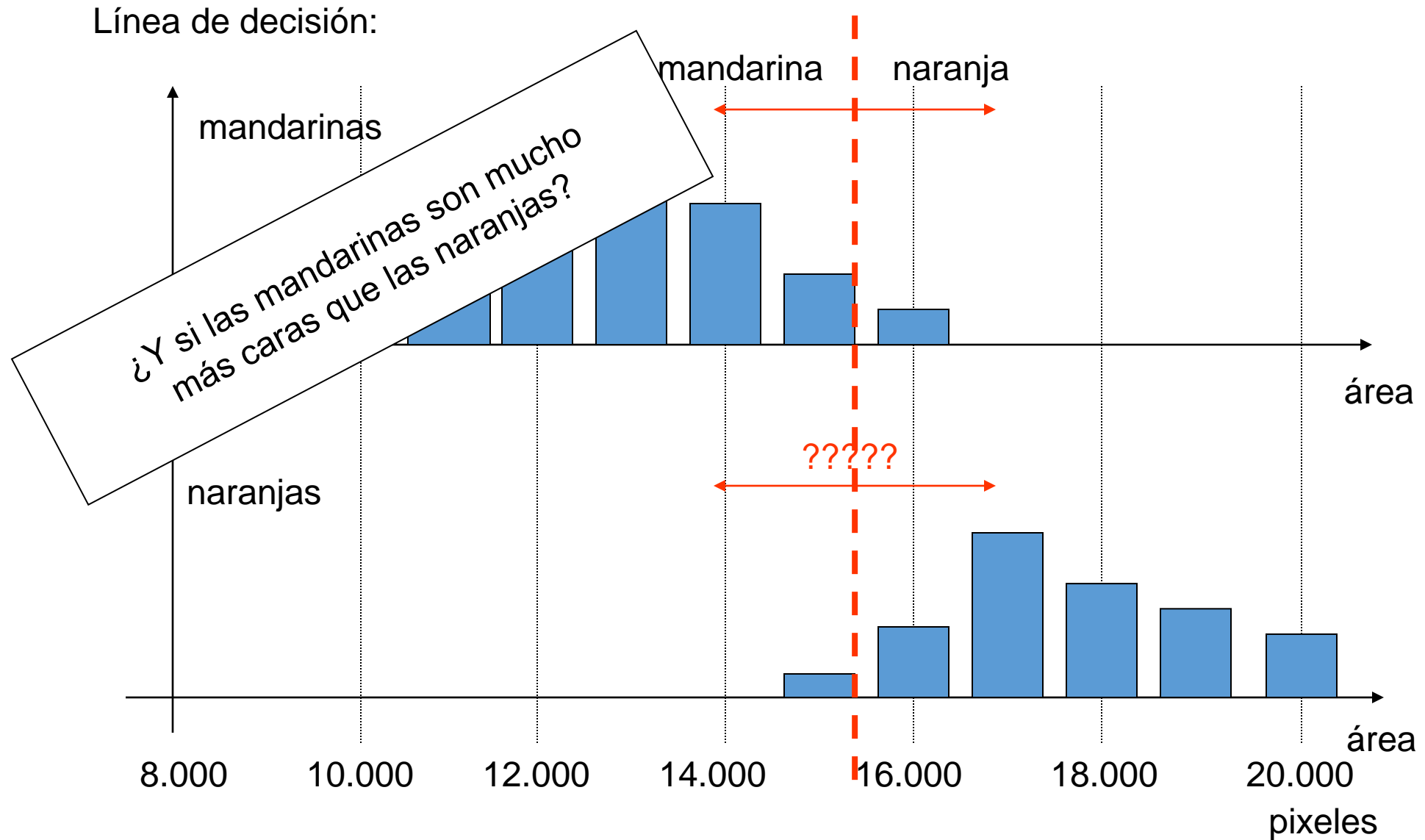
Línea de decisión:



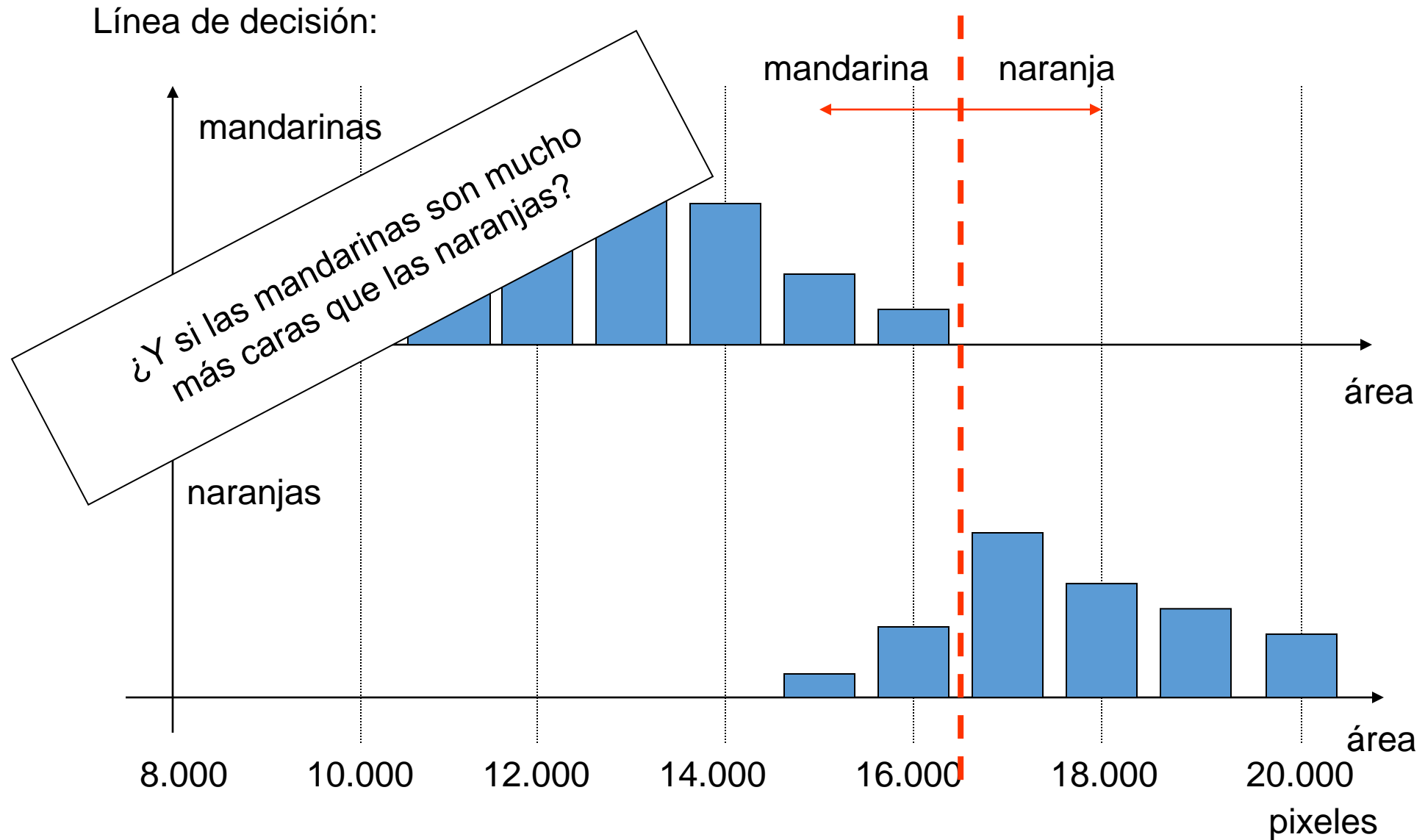


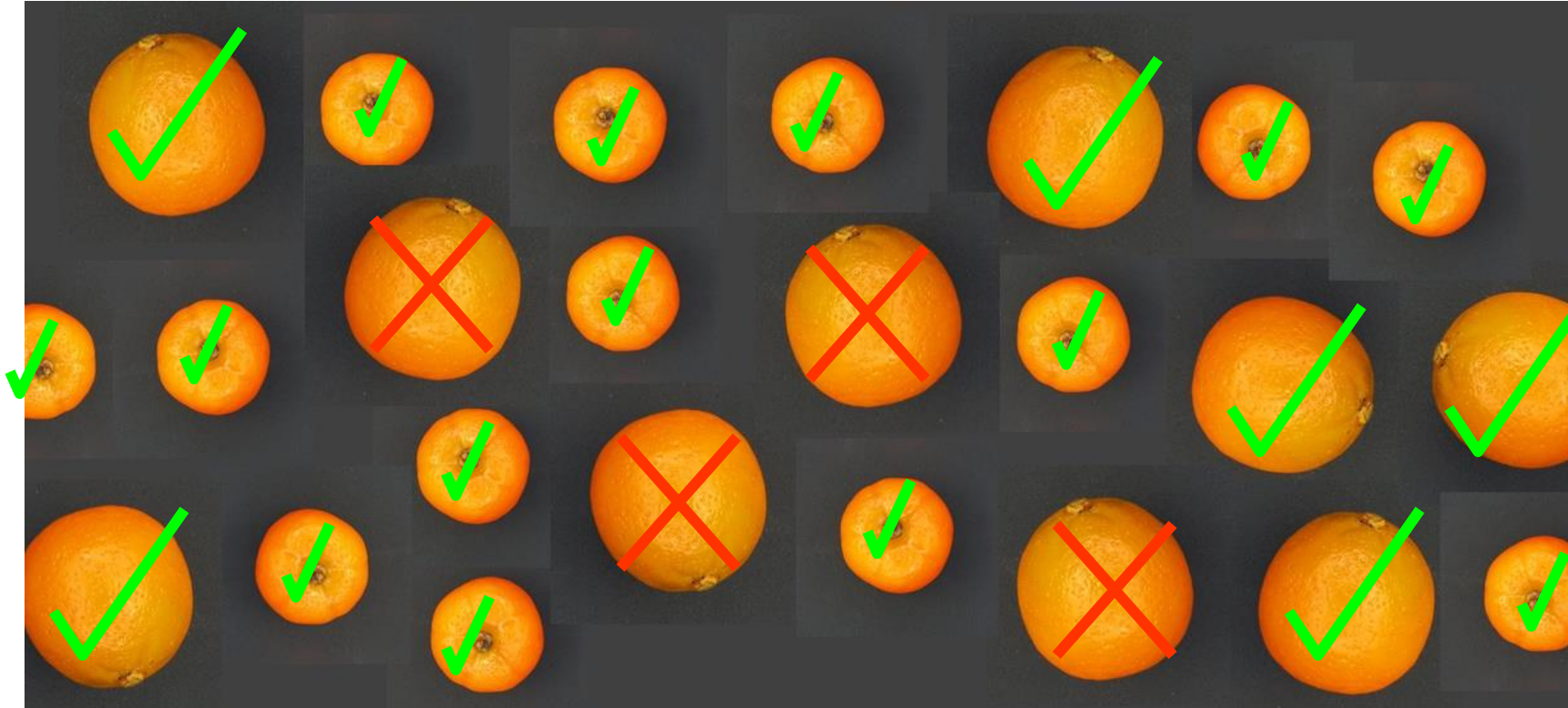
La separación es buena pero no es perfecta

Un ejemplo práctico



Un ejemplo práctico

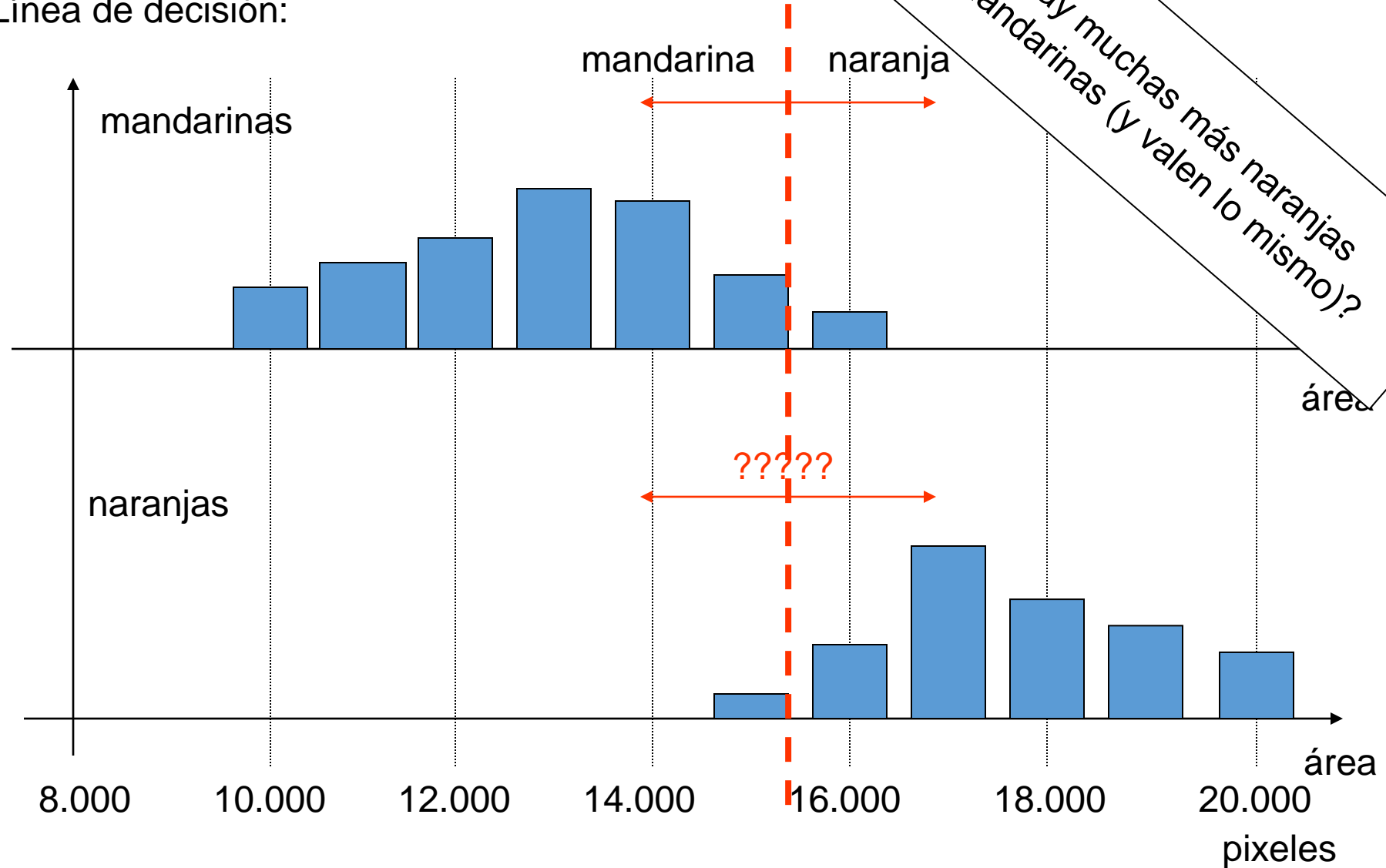




Todas las mandarinas son clasificadas perfectamente
... pero el costo es que hay varias naranjas mal clasificadas

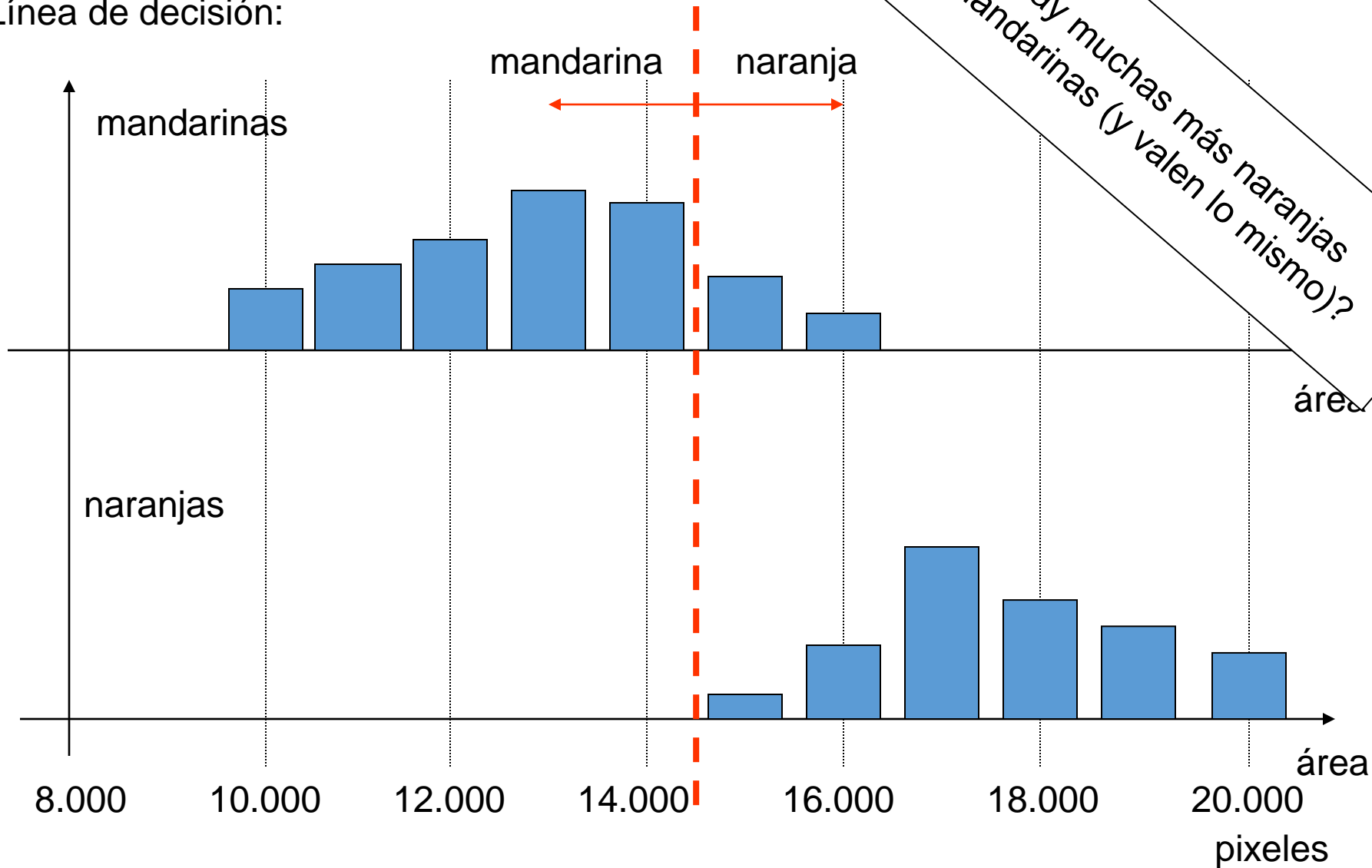
Un ejemplo práctico

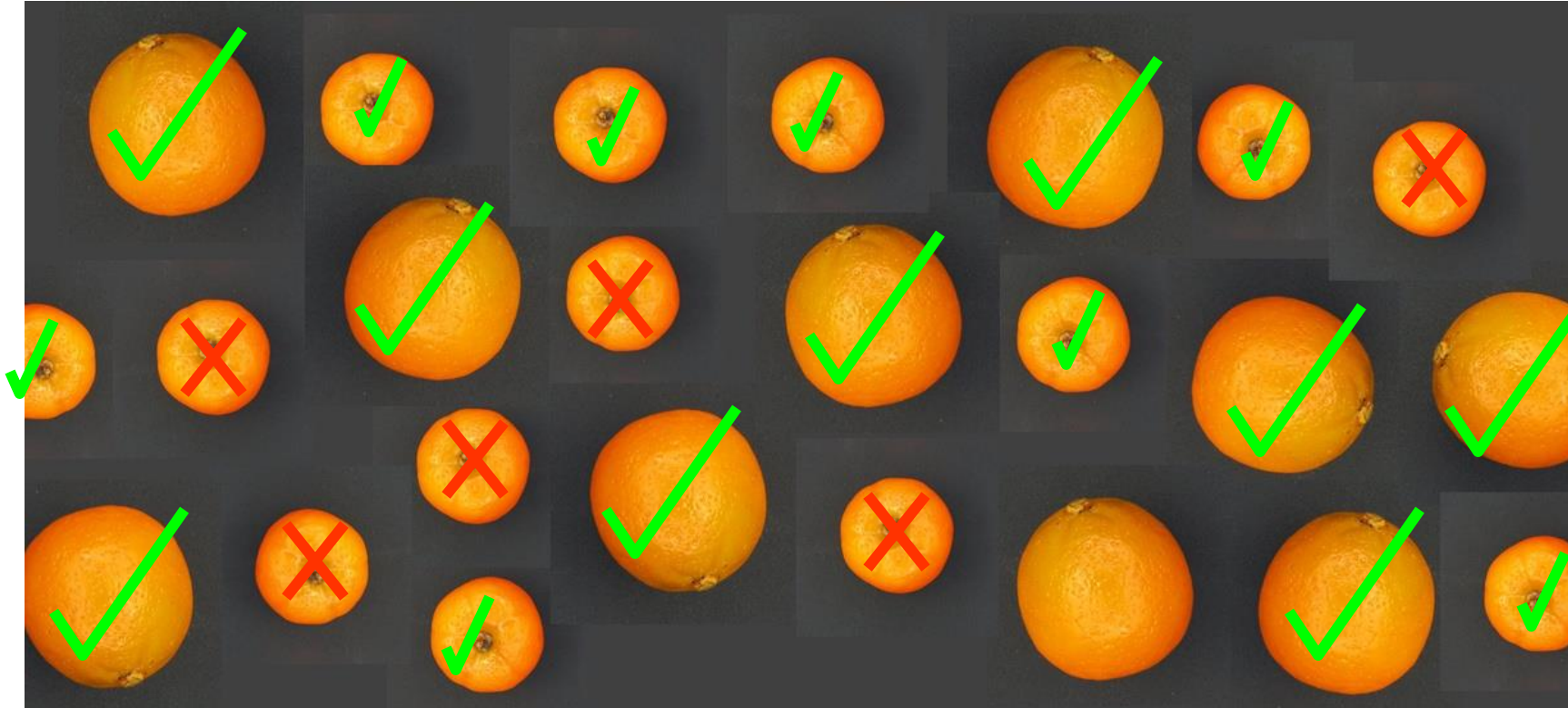
Línea de decisión:



Un ejemplo práctico

Línea de decisión:





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%

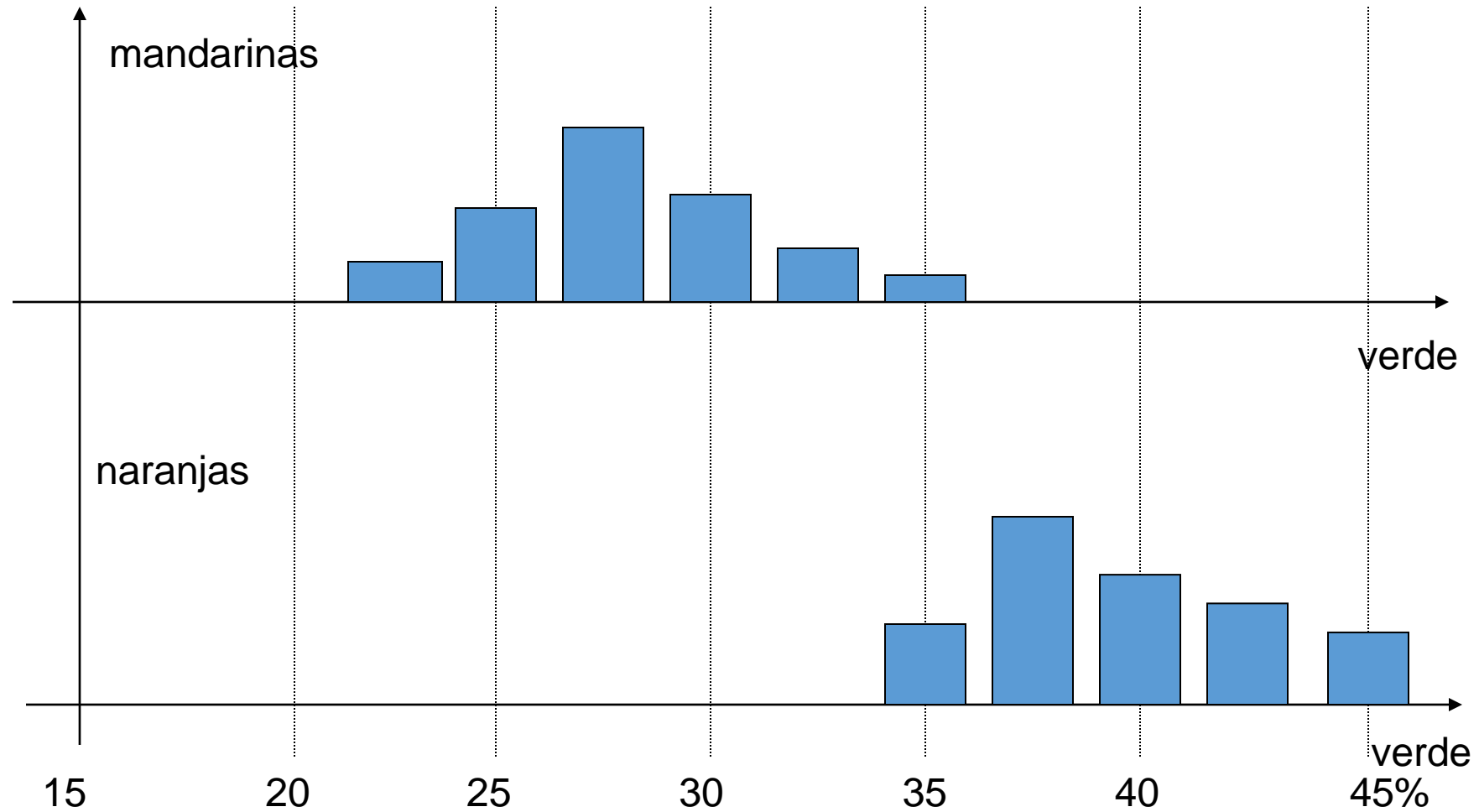


Verde = 46%

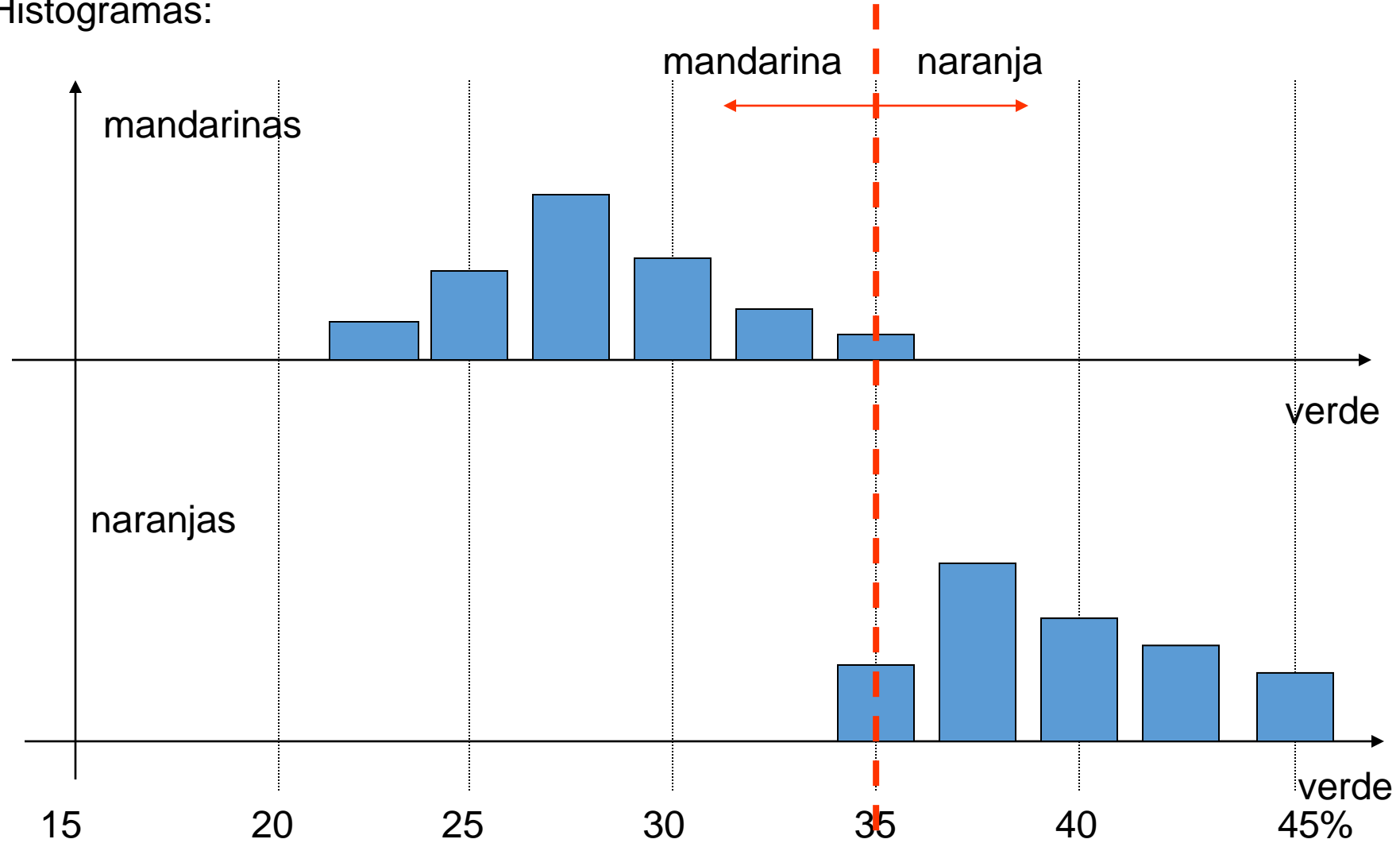
Recolección de información: Porcentaje de Verde

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

Histogramas:

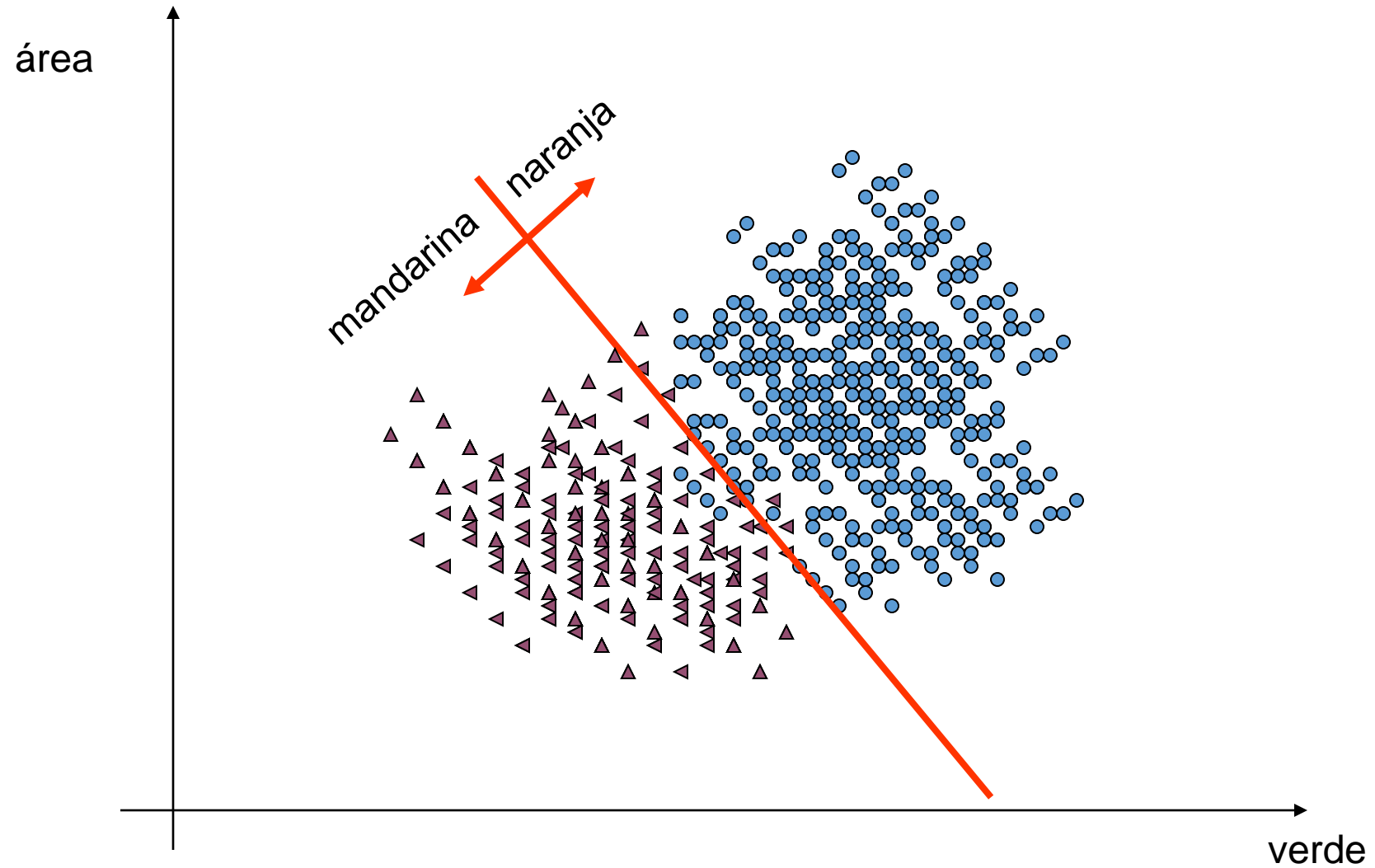


Histogramas:



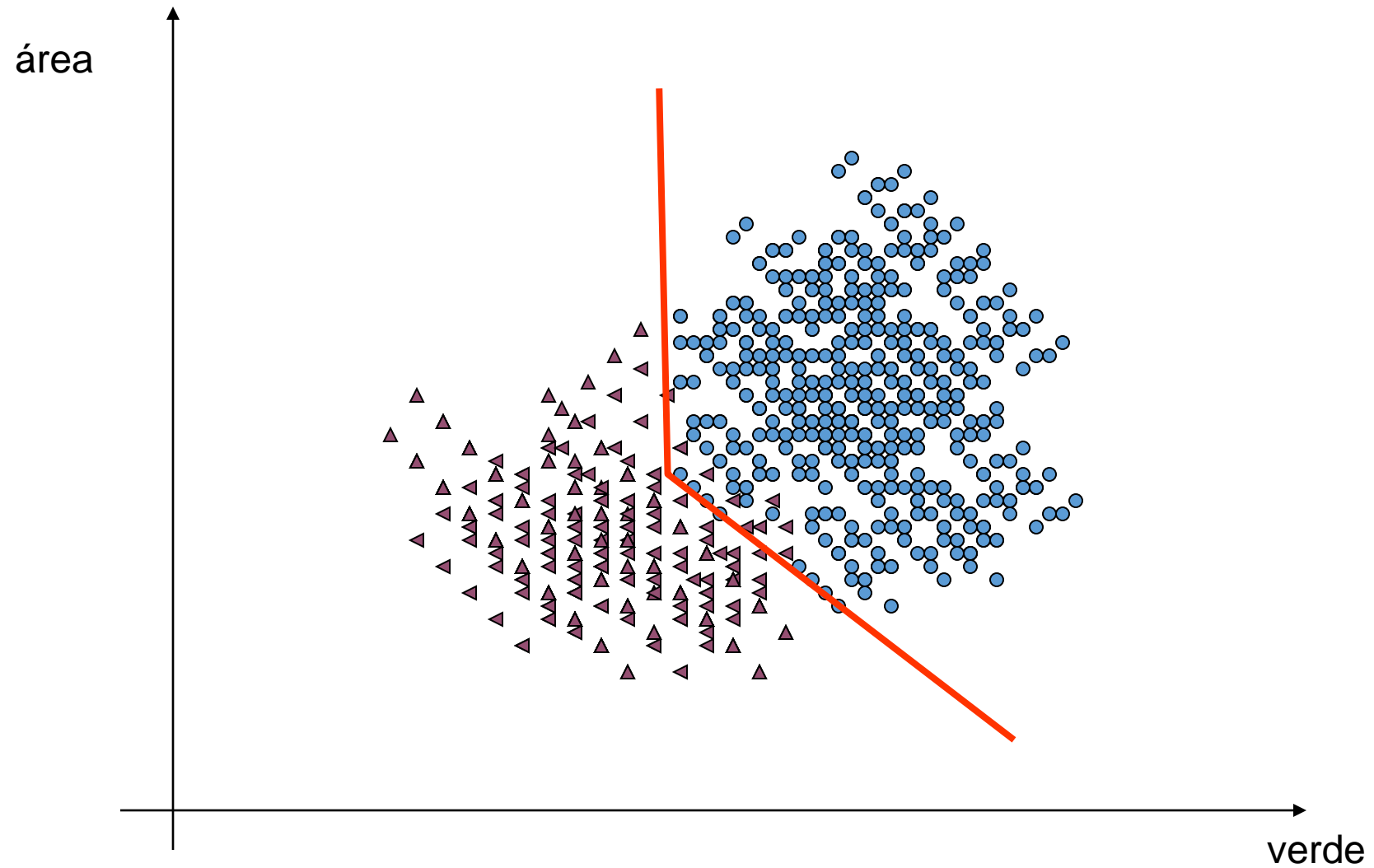
Un ejemplo práctico

Uso de dos características:



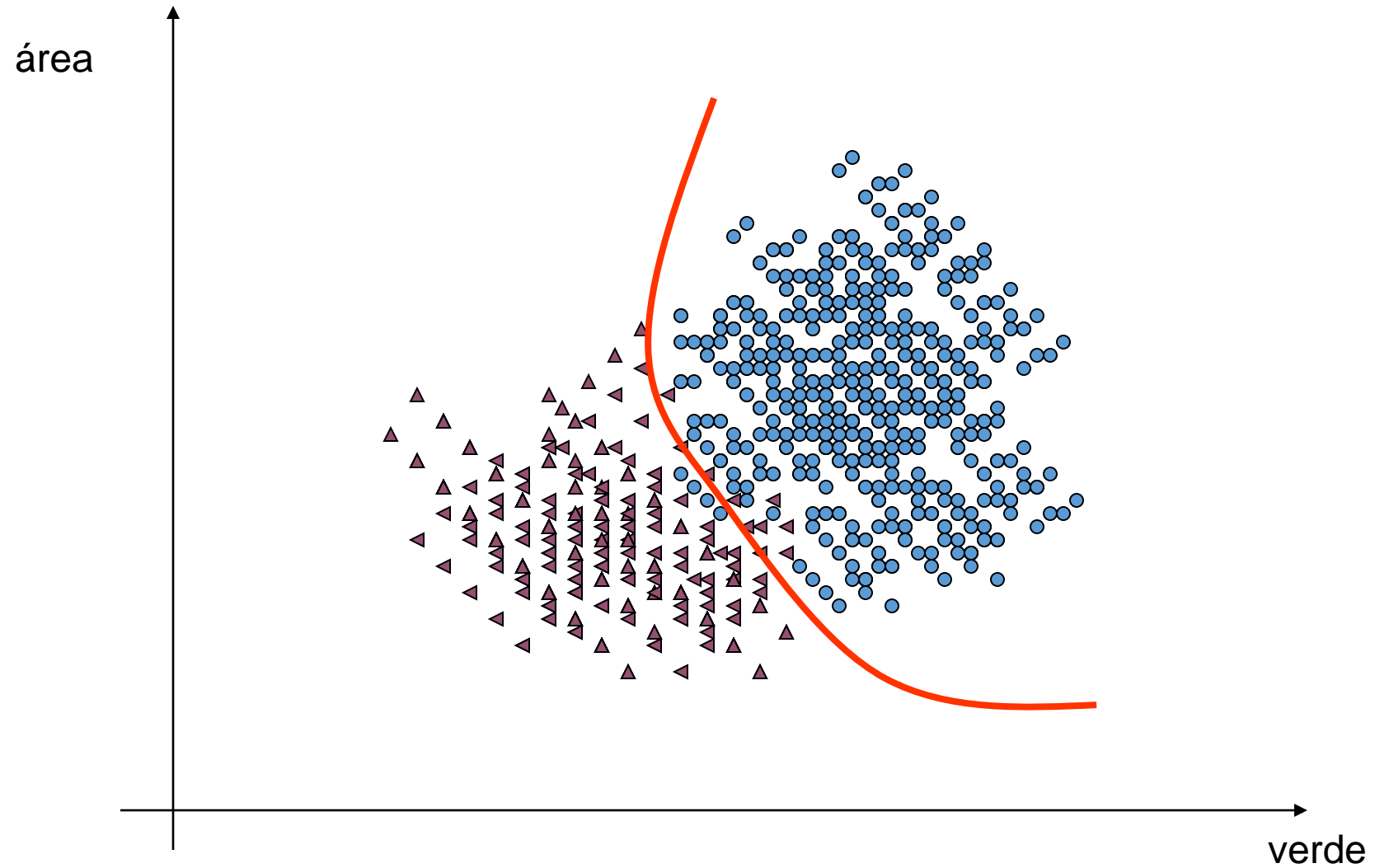
Un ejemplo práctico

Uso de dos características:



Un ejemplo práctico

Uso de dos características:



Un ejemplo práctico

Uso de dos características:

