

Aprendizaje Automático

Segundo Cuatrimestre de 2018



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Aprendizaje Automático

Segundo Cuatrimestre de 2018

Primera mitad de la clase de hoy:

- Cuestiones administrativas: horarios, docentes, web, etc.
- Objetivos, correlatividades, evaluación, bibliografía.
- Repaso superficial del contenido de toda la materia.

Segunda mitad: Primera clase teórica.

Aprendizaje Automático

Segundo Cuatrimestre de 2018

- Docentes:** Agustín Gravano (profesor)
Pablo Brusco (jefe de trabajos prácticos)
- Horario y lugar:** Jueves 13 a 17 horas. Aula 3 y laboratorio 6.
- Campus virtual:** www.dc.uba.ar/aa
Importante: Tienen que **matricularse** a la materia.
La comunicación será a través del campus virtual.
- Puntajes:** 3 puntos para Licenciatura y Doctorado (Computación)
- Correlativas:** Métodos Numéricos; Algoritmos y Estructuras de Datos 3

Aprendizaje Automático

Segundo Cuatrimestre de 2018

Modo de evaluación:

- 2 exámenes parciales.
- 2 trabajos prácticos grupales (tres integrantes).

Régimen de aprobación y promoción:

- Para **aprobar** la cursada deben aprobarse los 2 parciales y los 2 TPs.
- Se puede **promocionar** obteniendo nota 8 o superior en los 2 parciales, y además completando todos los ejercicios extras en los TPs.
- En caso de promoción, la **nota final** se determina globalmente, considerando todas las instancias de evaluación.
- Quienes obtengan nota inferior a 8 en algún parcial deben dar el **final**.
- Para quienes recuperen un parcial, la nota que cuenta es la del recuperatorio.
- La promoción es **opcional**: pueden optar por dar final para levantar la nota.

Aprendizaje Automático

Segundo Cuatrimestre de 2018

Grupos para los trabajos prácticos:

- Grupos de tres (3) integrantes.
- **Traer los grupos ya confirmados para la próxima clase.**
- Hagan lo posible para que:
 - 1) al menos un integrante tenga formación en programación;
 - 2) todos los integrantes estén en la misma situación (p.ej., les importa o no aprobar la materia; quieren intentar promocionar o no).

Aprendizaje Automático

Segundo Cuatrimestre de 2018

Bibliografía:

- Básica:

- James, Witten, Hastie & Tibshirani, "An Introduction to Statistical Learning with Applications in R", 6th ed, Springer, 2015.
- Mitchell, "Machine Learning", McGraw-Hill, 1997.
- Müller & Guido, "Introduction to Machine Learning with Python", O'Reilly, 2016.

- Avanzada:

- Hastie, Tibshirani & Friedman, "The Elements of Statistical Learning", 2nd ed, Springer, 2009.
- Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.
- Duda, Hart & Stork, "Pattern Classification", 2nd ed, Wiley, 2001.

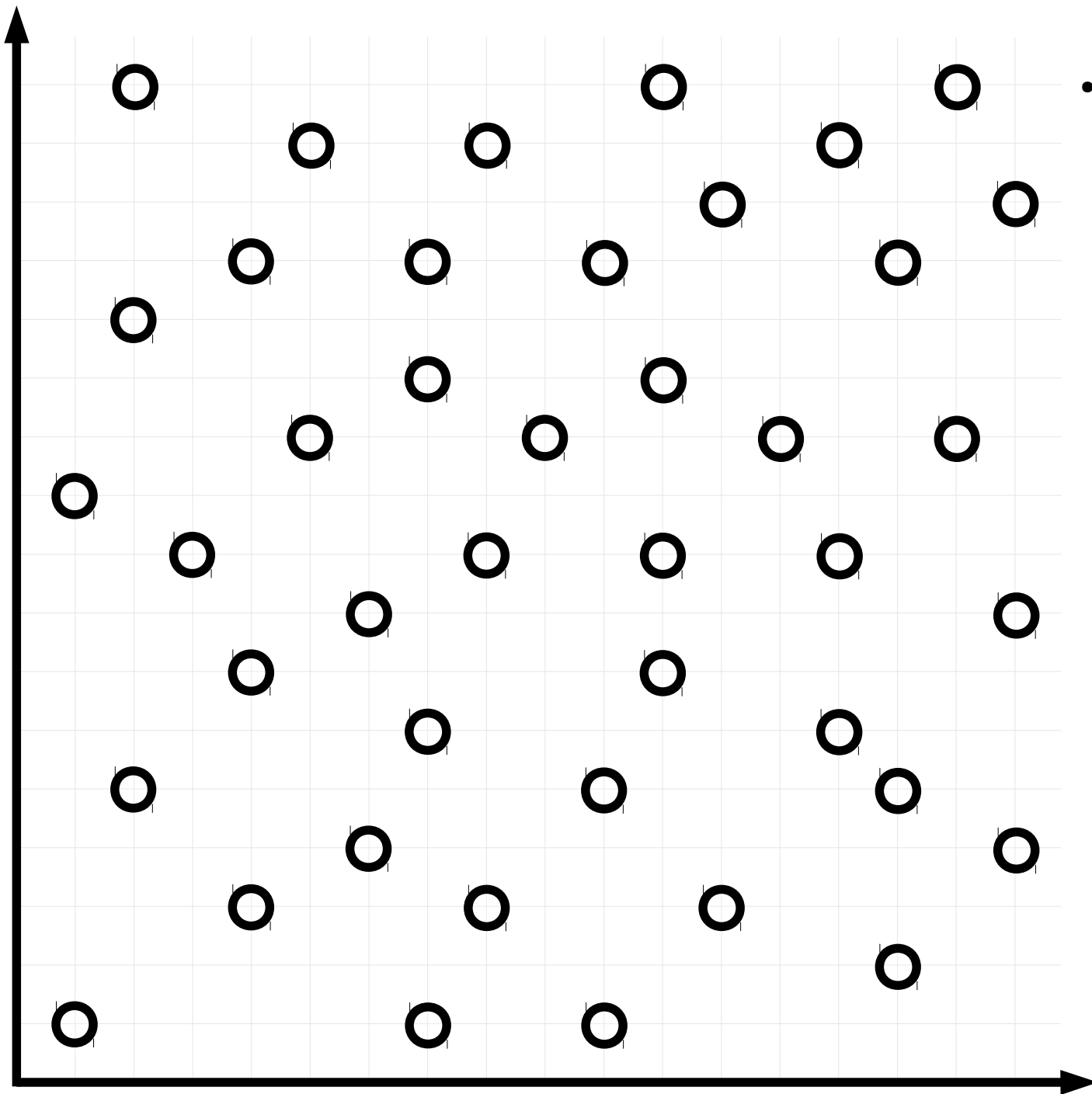
Aprendizaje Automático

Segundo Cuatrimestre de 2018

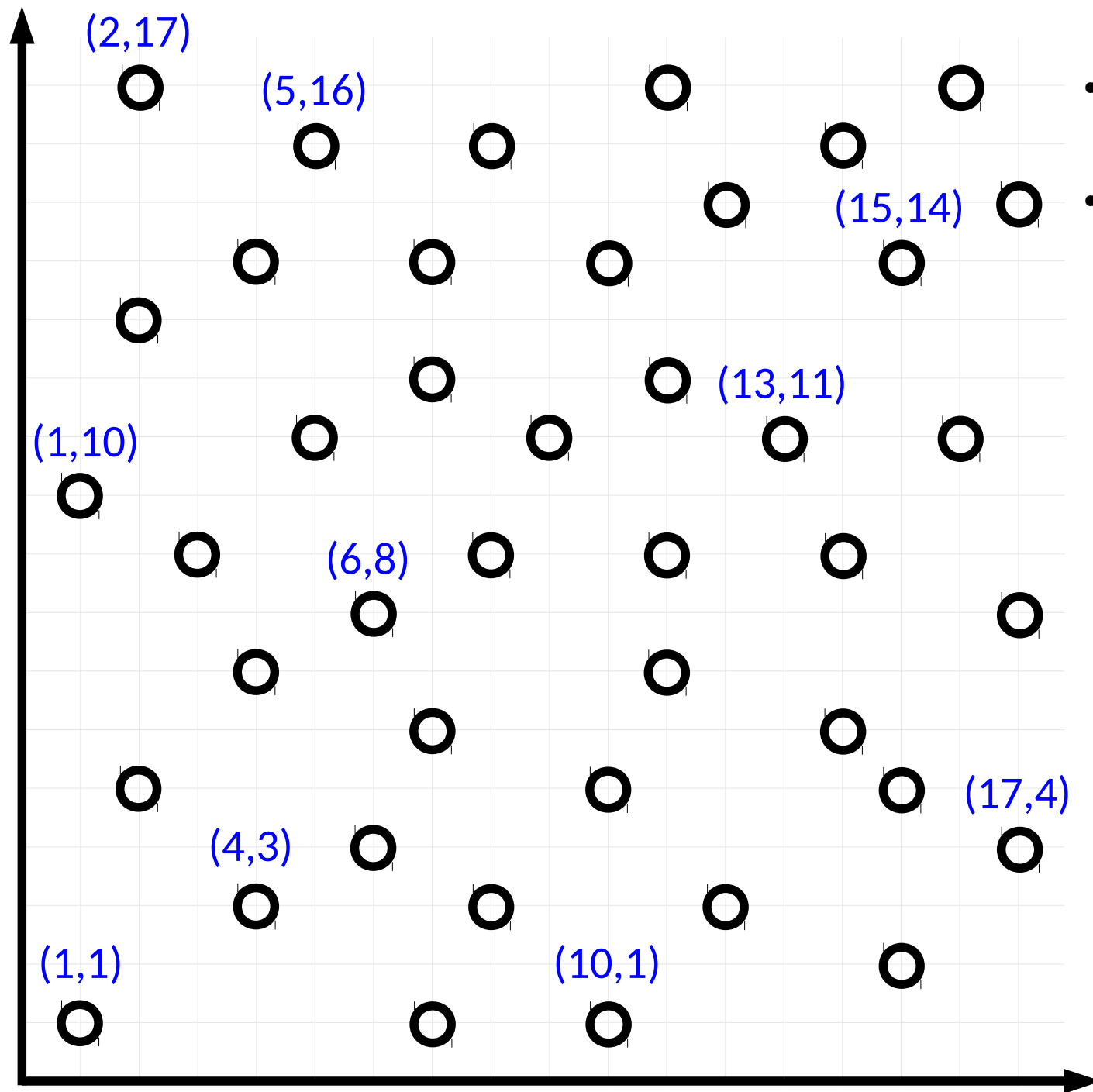
Objetivos:

El Aprendizaje Automático (o *Machine Learning*) es una rama de la Inteligencia Artificial que se dedica al estudio de los programas que aprenden a realizar una tarea en base a la experiencia.

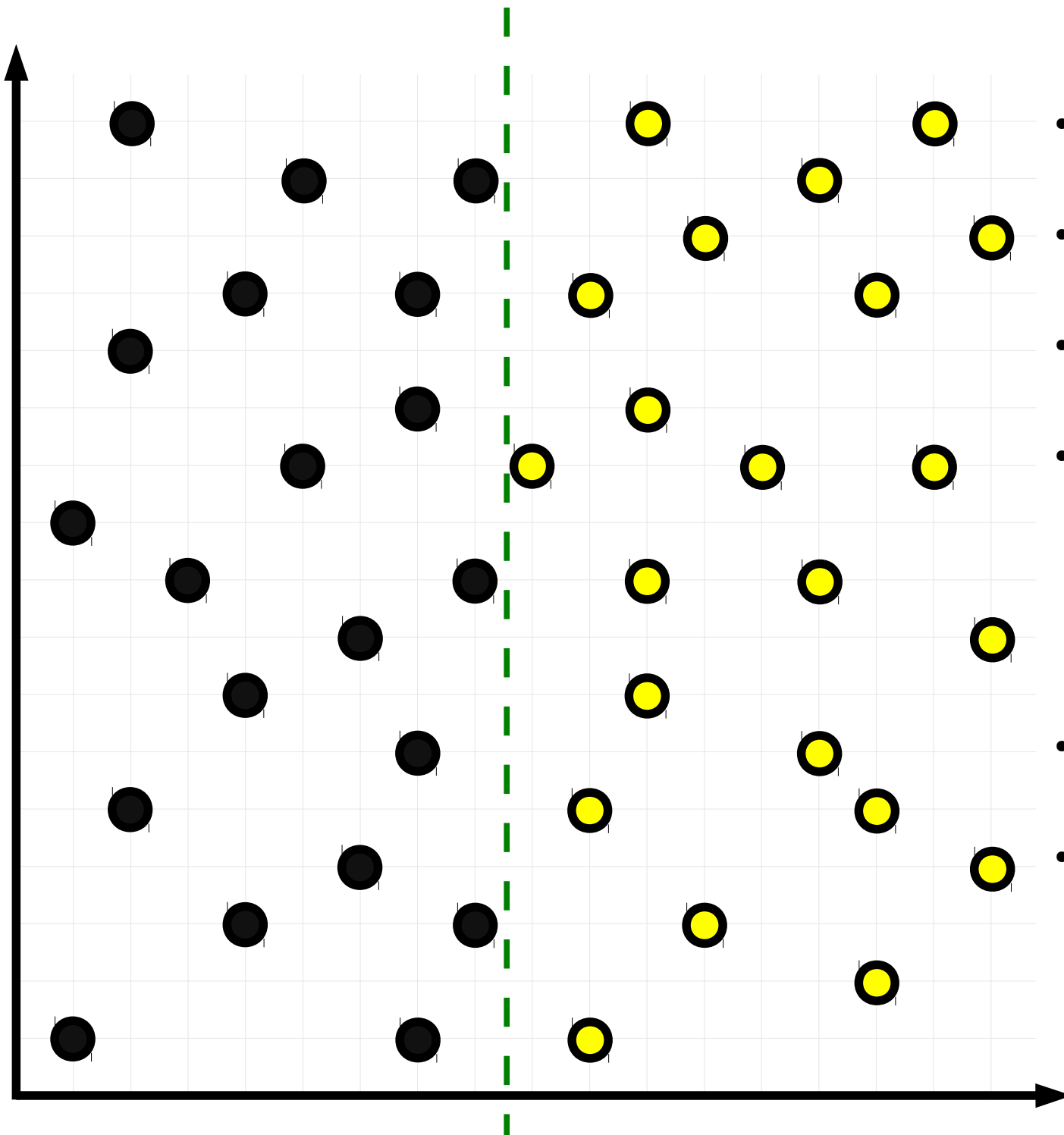
Esta materia consiste en una introducción abarcativa de las principales técnicas y aplicaciones del área, con un balance entre teoría y ejercitación práctica.



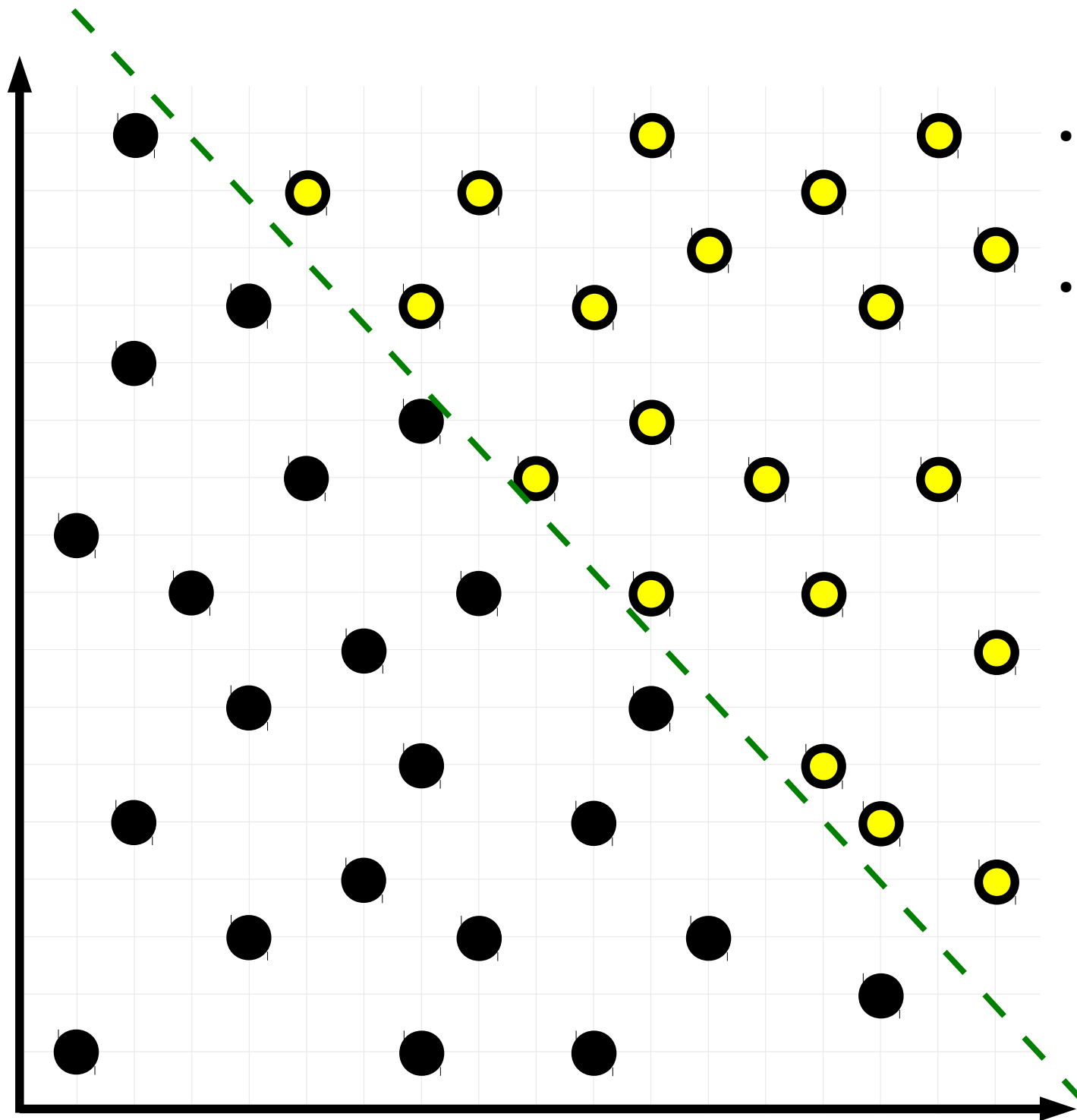
- Tenemos n puntos en el plano.



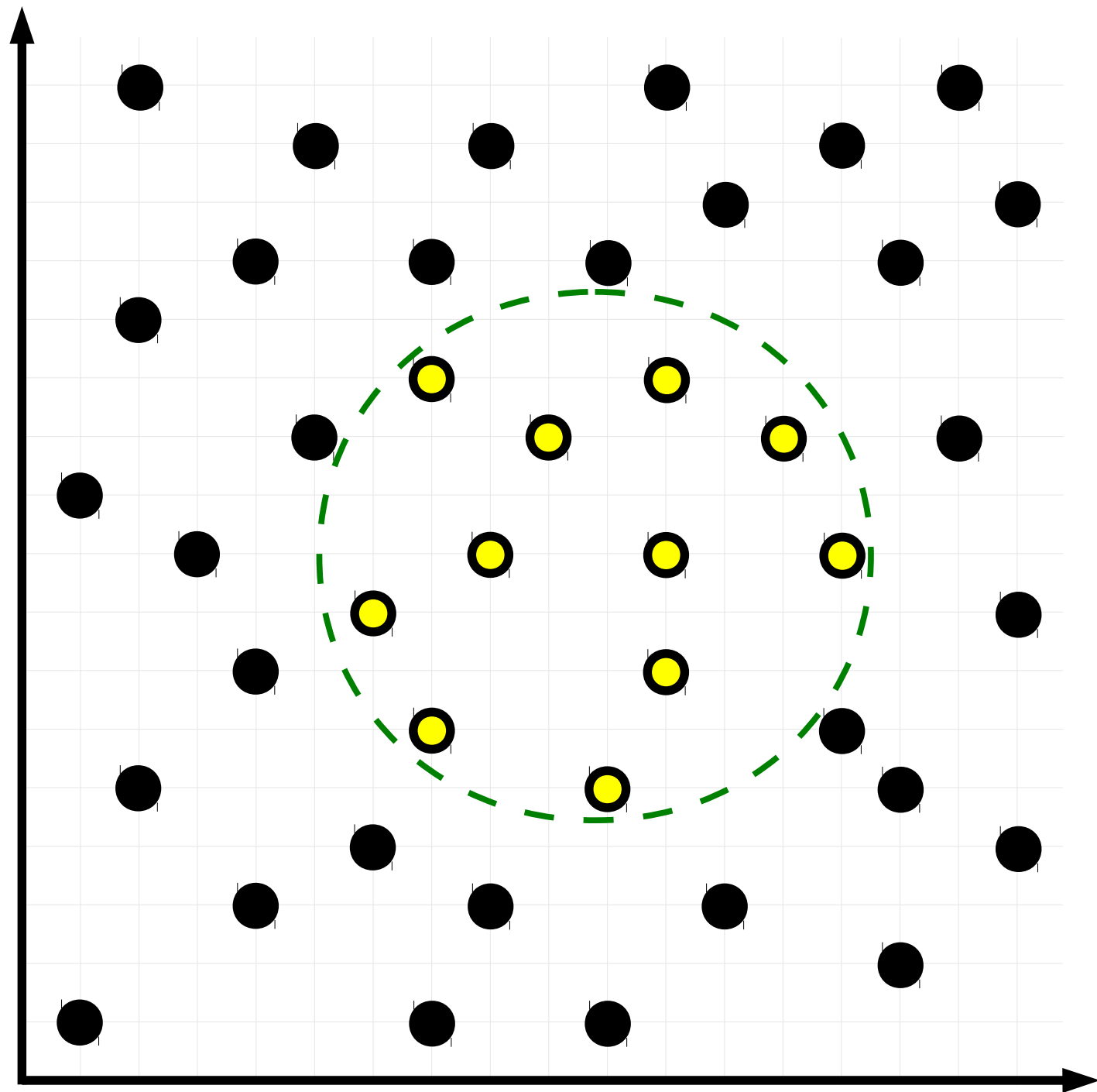
- Tenemos n puntos en el plano.
- Cada punto tiene dos coordenadas: (x, y) .

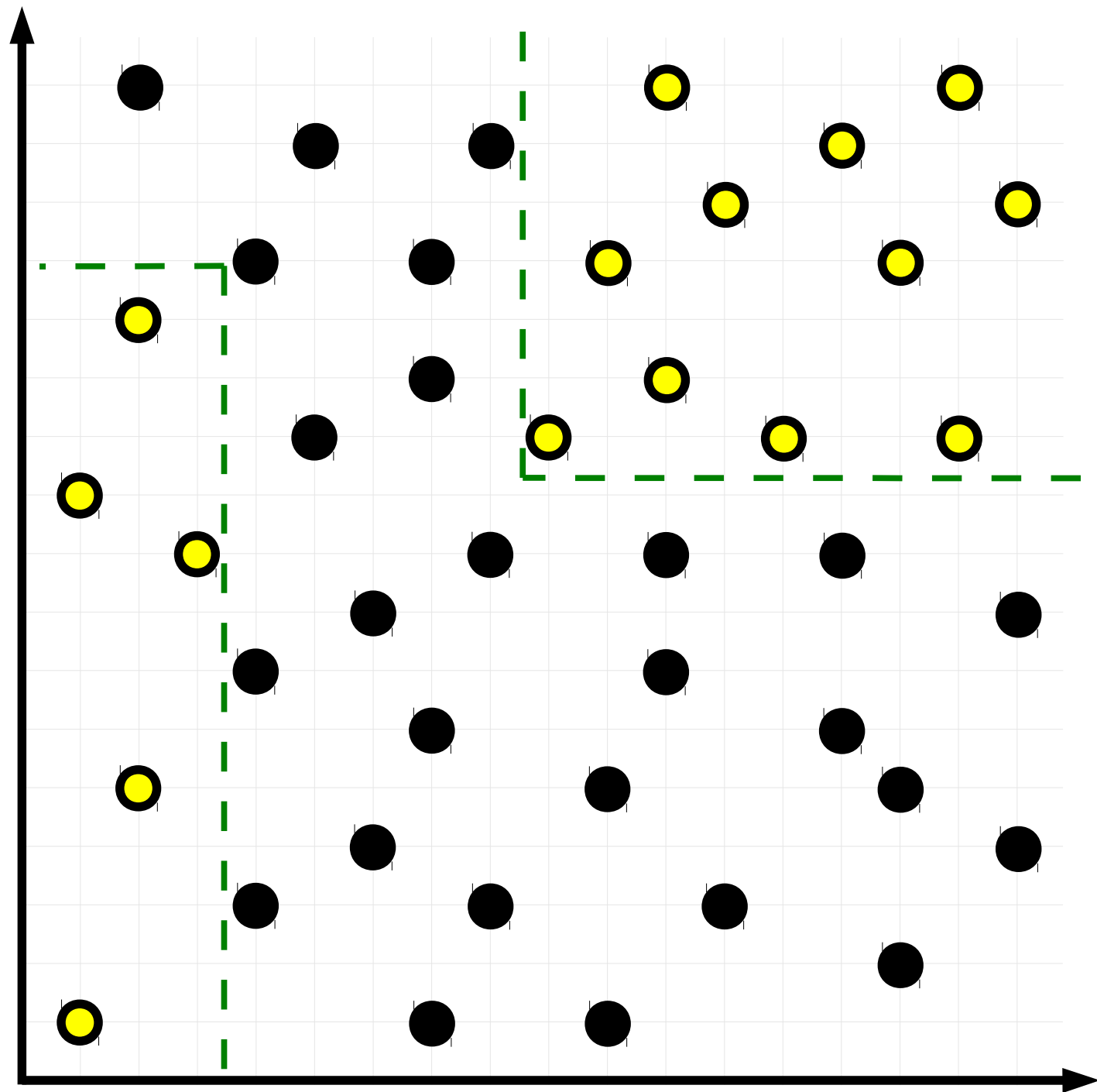


- Tenemos n puntos en el plano.
- Cada punto tiene dos coordenadas: (x, y) .
- Cada punto tiene un color: **negro** o **amarillo**.
- Queremos **aproximar** el criterio que determina el color de los puntos, ya sea:
 - para **predecir** el color de nuevos puntos, o
 - para **describir** el mundo.
- Existe una función:
 $f(x, y) \rightarrow \text{color}$
- La aproximamos mediante:
$$\hat{f}(x, y) = \begin{cases} \text{amarillo} & \text{si } x > 8 \\ \text{negro} & \text{si no} \end{cases}$$



- $\hat{f}(x, y) =$
 colorizq si $y > m x + b$
 el otro color, si no
- *colorizq*, *m* y *b* son
 parámetros del modelo que
 deben ajustarse a los datos.





Humanos vs. Máquinas

- Los humanos somos buenos encontrando (y programando) estas reglas en 2D.
- Pero, ¿qué pasa si los puntos tienen miles de coordenadas?
- Ejemplo: Detección de caras.



- Los humanos somos muy buenos detectando (y reconociendo) caras.
- Pero ¿podemos **programar estas funciones?**

Aprendizaje Automático

Un programa **aprende** una tarea, si su performance mejora con la experiencia.

Tenemos que definir:

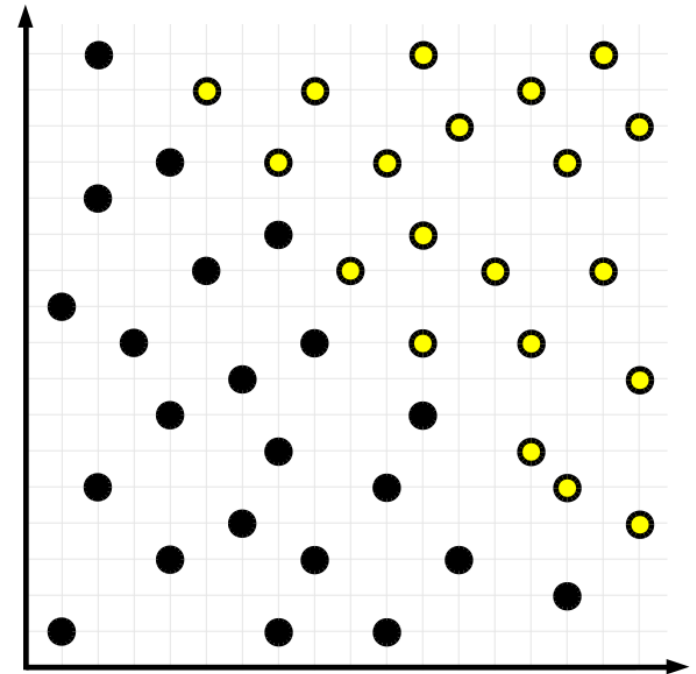
- **Tarea**
- **Medida de performance**
- **Experiencia**

Aprendizaje Automático

Un programa **aprende** una tarea, si su performance mejora con la experiencia.

Ejemplo:

- **Tarea:** Predecir el color de un punto.
- **Medida de performance:** % puntos coloreados correctamente.
- **Experiencia:** Base de datos de puntos con su color correspondiente.

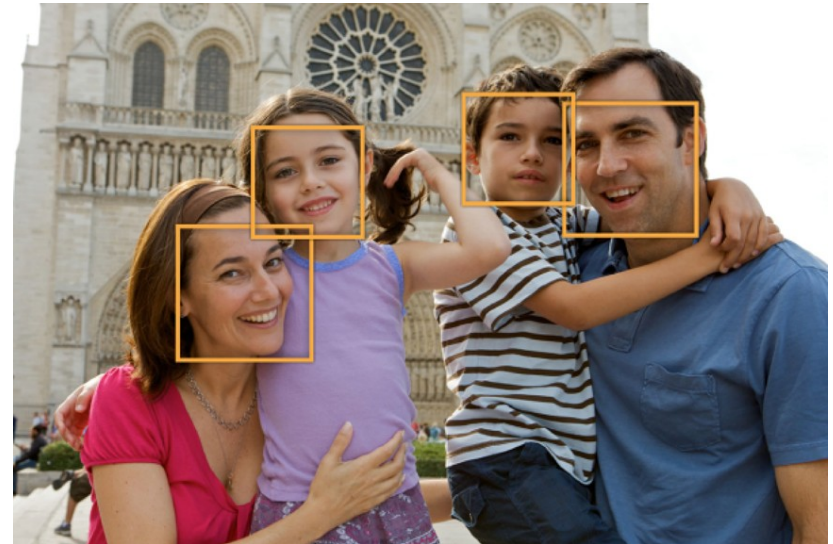


Aprendizaje Automático

Un programa **aprende** una tarea, si su performance mejora con la experiencia.

Ejemplo:

- **Tarea:** Detectar caras en una imagen.
- **Medida de performance:** % caras detectadas correctamente.
- **Experiencia:** Base de datos de imágenes con las caras marcadas.

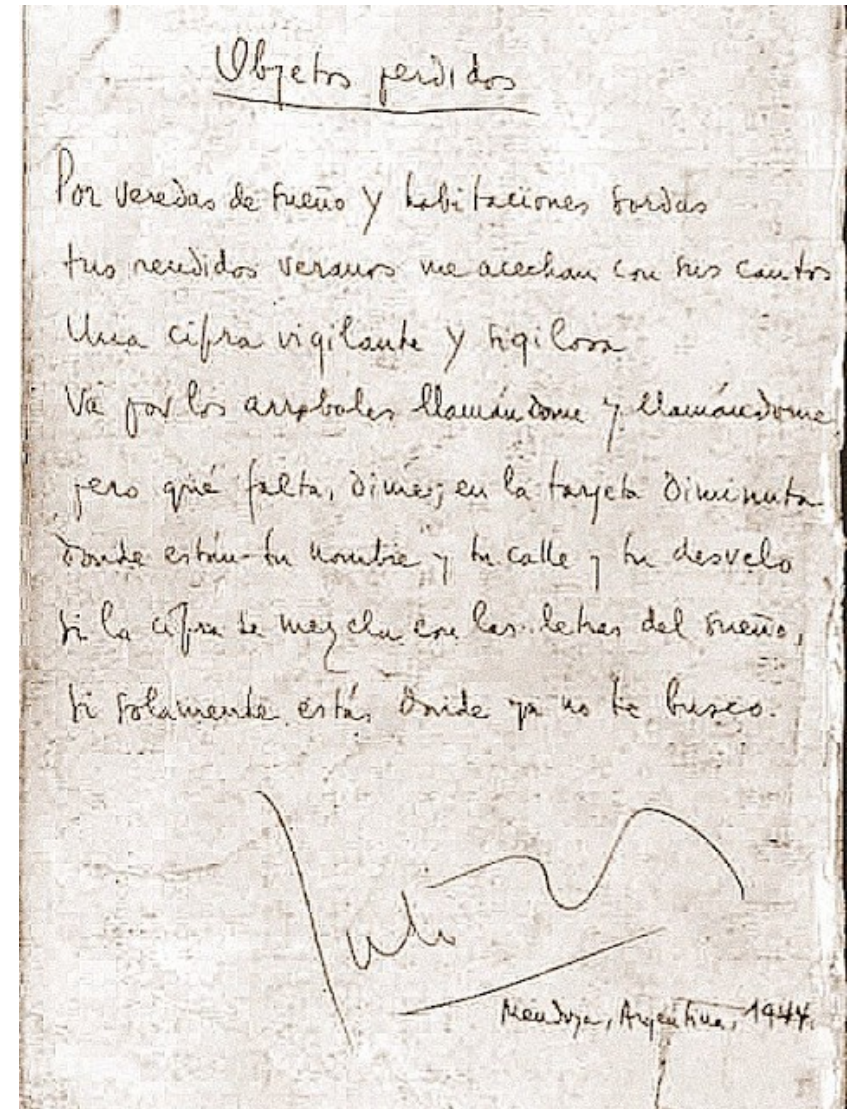


Aprendizaje Automático

Un programa **aprende** una tarea, si su performance mejora con la experiencia.

Ejemplo:

- **Tarea:** Reconocer textos manuscritos.
- **Medida de performance:** % palabras reconocidas correctamente.
- **Experiencia:** Base de datos de palabras manuscritas con sus transcripciones correctas.



Aprendizaje Automático

Un programa **aprende** una tarea, si su performance mejora con la experiencia.

Ejemplo:

- **Tarea:** Manejar un auto por la calle usando sensores visuales.
- **Medida de performance:** Distancia promedio recorrida hasta cometer un error.
- **Experiencia:** Secuencia de *mediciones* (ej: imágenes, acelerómetros) y *acciones* (acelerar, frenar, doblar, etc.) grabados mientras conduce un ser humano.





Aprendizaje Automático

Más ejemplos:

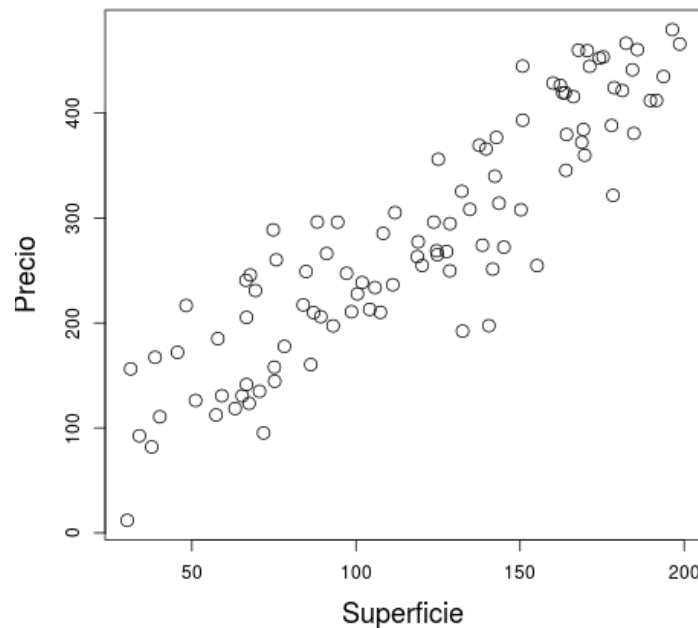
- Clasificación de mensajes (ej: spam):
 - GMail, Hotmail, Yahoo! mail
- Reconocimiento del habla:
 - Siri, Cortana, Google Now, Amazon Echo
- Sistemas de recomendación:
 - Spotify, Facebook, Netflix, Amazon, Mercado Libre, Google
- Predicción de tiempo de viaje, camino óptimo:
 - Waze, Uber, Google Maps, Despegar
- Detección de fraude:
 - PayPal, Mercado Libre, bancos
- Publicidad online:
 - Google Ads, Jampp
- ...

Empresas más grandes del mundo:

1997	2007	2017
		
		
		
		
		

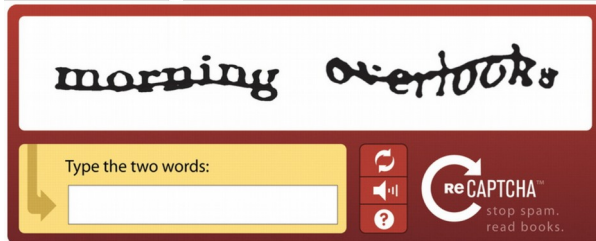
Aprendizaje Supervisado

- Tipos de aprendizaje supervisado:
 - **Clasificación**: Cada instancia pertenece a una **clase**.
 - SpamFilter: mensaje \rightarrow {spam, no-spam}
 - **Regresión**: Cada instancia tiene un **valor numérico**.
 - Tasador: propiedad \rightarrow precio (en miles de USD)



Aprendizaje Supervisado

- Los datos de entrenamiento están **anotados** con la respuesta correcta.
 - Típicamente: anotación manual, costosa!
 - Amazon Mechanical Turk
 - reCAPTCHA




Provide 3 tags for this image.

Instructions:

- You must provide 3 tags for this image.
- Each tag must be a single word
- No tag can be longer than 25 characters
- The tags must describe the image, the contents of the image, or some relevant context.

Image:



Tag 1:

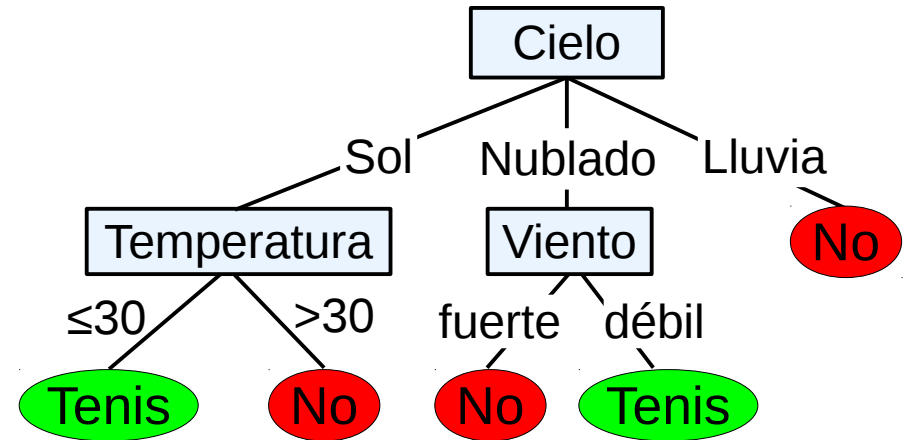
Tag 2:

Tag 3:

- ¡Que los usuarios hagan el trabajo gratis! Ejemplo: 

Algoritmos de Aprendizaje Supervisado

- Árboles de decisión



- Reglas

IF (Cielo=Sol \wedge Temperatura>30) THEN Tenis=No

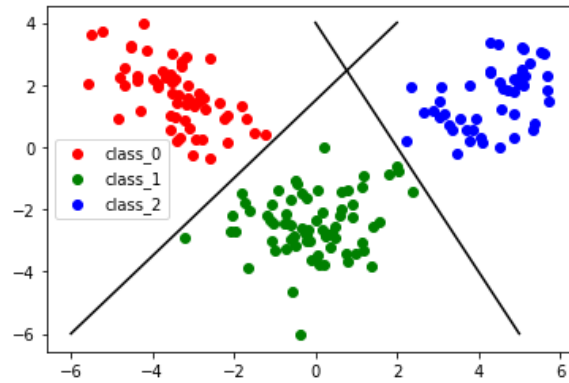
IF (Cielo=Nublado \wedge Viento=Débil) THEN Tenis=Sí

IF (Cielo=Lluvia) THEN Tenis=No

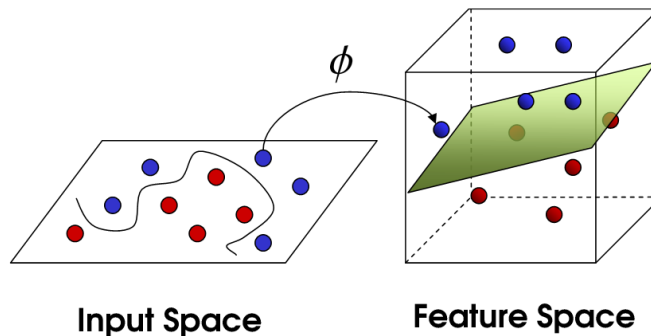
- Vecinos más cercanos (KNN)
- Naive Bayes

Algoritmos de Aprendizaje Supervisado

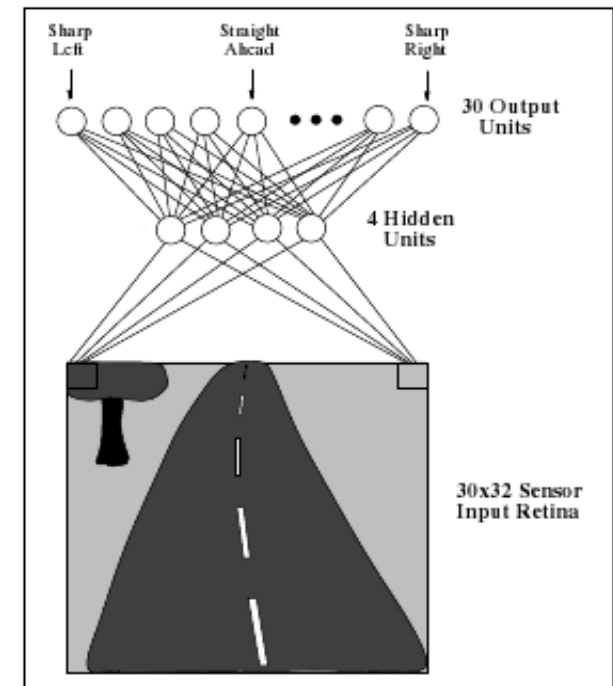
- Linear Discriminant Analysis



- Support Vector Machines



- Redes Neuronales



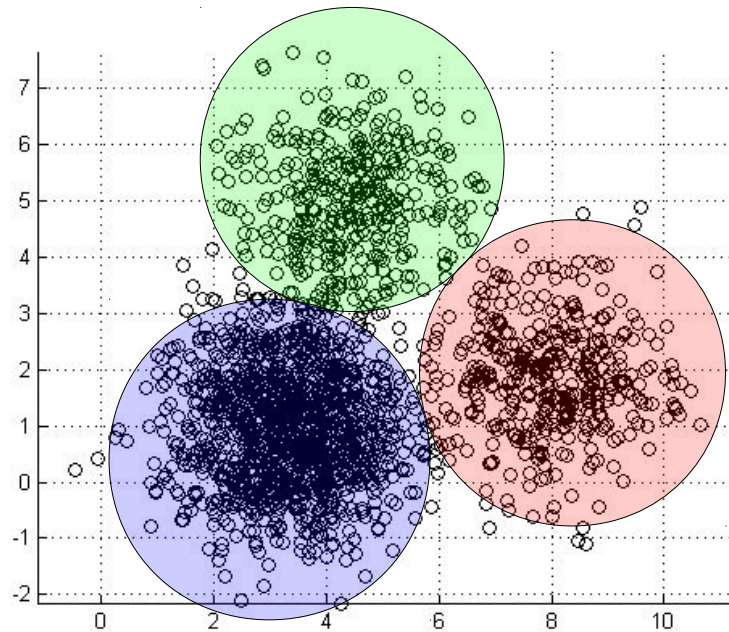
- ...

Esquema General de Aprendizaje Supervisado

- Definición de la tarea de aprendizaje.
 - Instancias, clases, medidas de performance.
- Recolección y preparación de **datos**.
 - Cantidad vs. calidad de datos. Ruido.
 - Extracción de atributos.
- Experimentación
 - Selección/transformación de los atributos.
 - Elección de algoritmos.
 - Entrenamiento y validación de modelos.
- Evaluación del modelo final sobre datos frescos.

Aprendizaje No Supervisado

- Los datos de entrenamiento **no están anotados**.
 - Clustering, para encontrar patrones ocultos.



- Reducción de dimensionalidad y transformación de los datos, para poder entenderlos mejor (ej: PCA, MDS).

Aprendizaje Automático

- Aprendizaje **supervisado**
 - Los datos de entrenamiento están anotados.
 - Clasificación. Regresión.
- Aprendizaje **no supervisado**
 - Los datos de entrenamiento no están anotados.
 - Clustering. Transformación de datos.
- Aprendizaje **por refuerzos**
 - Aprendizaje gradual, en base a premios y castigos.

Aprendizaje **por Refuerzos**

- Conjunto de **estados** que definen el medio.
- Conjunto de **acciones** que el agente puede realizar.
- Reglas de **transiciones** entre estados.
- Reglas que asignan una **recompensa** a cada transición.
- Reglas que determinan qué **observa** el agente.
- Ejemplos: control de robots, programa de ascensores, ruteo de paquetes, juego del Go.



Temas hasta el primer parcial

- Aprendizaje de conceptos.
- Árboles de decisión.
- Evaluación de modelos.
- Más clasificadores: KNN, Naive Bayes, SVM, etc.
- Sesgo y varianza. Ensamblados.
- Preprocesamiento de datos.
- Reducción de dimensionalidad. Transformación de atributos.
- Clustering: k-means (algoritmo EM), BDSCAN, etc.
- Primer parcial.
- **Después:** Regresión lineal simple, múltiple, logística. Modelos generativos y discriminativos. Redes Neuronales. Aprendizaje por refuerzos. Aplicaciones.