



# MAESTRÍA EN CIENCIA DE DATOS

## Machine Learning y Deep Learning Aplicado

**Dr. José Eduardo Ochoa Luna**



Universidad Católica  
**San Pablo**

Departamento de Ciencia  
de la Computación

**6 Marzo 2024**

# Agenda

- Estructura del curso
- Revisión Aprendizaje de máquina (computacional)

# Objetivos del curso

- Conocer los diversos algoritmos de Machine Learning y Deep Learning que permitan construir modelos predictivos
- Desarrollar proyectos de análisis de datos usando Machine Learning
- Desarrollar proyectos de análisis de datos usando Deep Learning

# Bibliografía

- Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, 2nd Edition, Wiley, 2000.
- Christopher M. Bishop, Pattern Recognition and Machine Learning. 2007.
- Tom Mitchell, Machine Learning. 1997.
- **Aurélien Géron. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition. 2019**

# Bibliografía

- **Ian Goodfellow, Yoshua Bengio and Aaron Courville. Deep Learning. Mit Press. 2016.**
- Yann LeCun, Yoshua Bengio and Geoffrey Hinton. Deep learning. Nature. 2015.
- Dive into Deep Learning (<https://d2l.ai/>)
- Tunstall et al. Natural Language Processing with Transformers. O'reilly. 2022

# Herramientas

- Jupyter
- Scikit-learn
- Tensorflow / PyTorch
- Hugging face
- Large Language Models (LLMs): OpenAI, Llama2, Mistral, Gemini, Gemma,....
- Langchain, LlamaIndex

# Contenido

Módulo	Profesores	Fecha
Introducción al curso	Dr. José Ochoa	06/03/2024
Regresión	MSc. Rosa Paccotacya	13/03/2024
Clasificación	MSc. Rosa Paccotacya	20/03/2024
Clustering	MSc. Rosa Paccotacya	03/04/2024
Redes Neuronales I	Dr. José Ochoa	17/04/2024
Redes Neuronales II	Dr. José Ochoa	24/04/2024

# Contenido

Módulo	Profesores	Fecha
Redes Neuronales Recurrentes I	Dr. José Ochoa	08/05/2024
Redes Neuronales Recurrentes II	Dr. José Ochoa	15/05/2024
Series de tiempo con RNNs	Dr. José Ochoa	22/05/2024
Mantenimiento Predictivo con RNNs	Dr. José Ochoa	29/05/2024
Redes Neuronales Convolucionales I	MSc. José Chávez	05/06/2024
Redes Neuronales Convolucionales II	MSc. José Chávez	12/06/2024



# Contenido

Módulo	Profesores	Fecha
GANs	MSc. José Chávez	19/06/2024
Transformers I	Dr. José Ochoa	26/06/2024
Transformers II	Dr. José Ochoa	03/07/2024
IA Generativa	Dr. José Ochoa	10/07/2024

# Evaluación

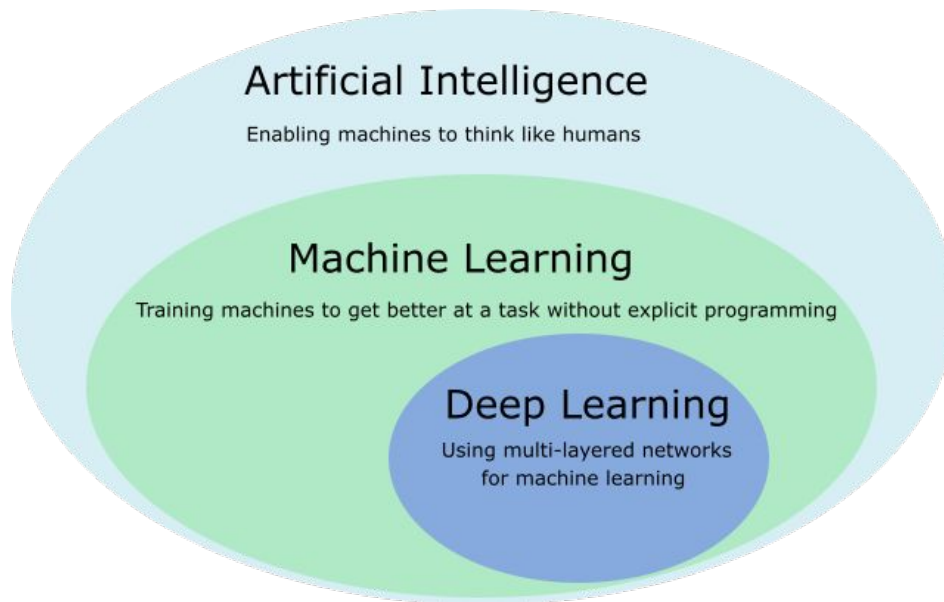
- Basado en ejercicios
- Basado en proyectos:  
Proyectos Machine Learning tradicional  
Proyecto de Deep Learning

# Revisión ML, Aprendizaje

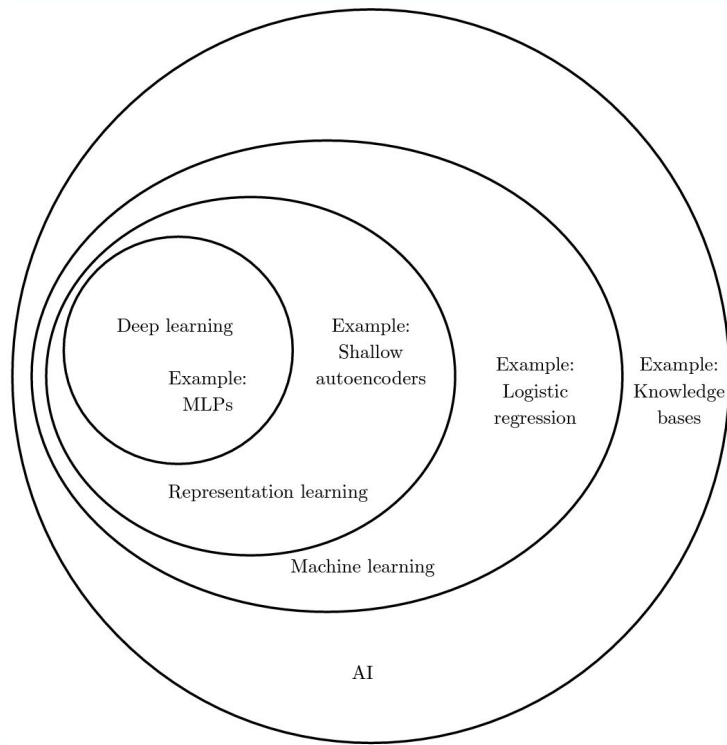
- Machine Learning
- Aprendizaje
- Regresión / Clasificación / Clustering

# Inteligencia Artificial y Machine Learning

## Computer Science



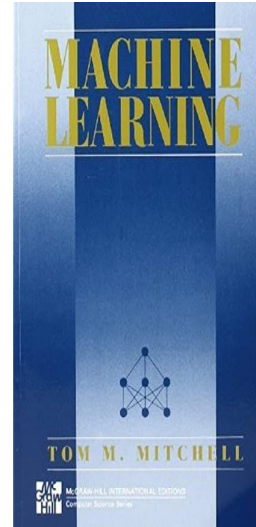
# Machine Learning y Deep Learning



# ¿Qué es Machine Learning ?

Un programa se dice que aprende de la experiencia  $E$  en relación a alguna tarea  $T$  y una medida de desempeño  $P$ , si su desempeño en  $T$ , medido por  $P$  mejora con la experiencia  $E$ .

Tom Mitchell, 1997



# Ejemplo Machine Learning

El filtro de Spam es un programa ML, que puede aprender a identificar spam a partir de ejemplos de emails spam (identificado por usuarios) y ejemplo de emails regulares (no spam)



# Ejemplo Machine Learning

La tarea T es identificar spam en nuevos emails, la experiencia son los datos de entrenamiento y el desempeño P, podría ser el ratio de emails correctamente clasificados (*accuracy*), usado en tareas de clasificación.



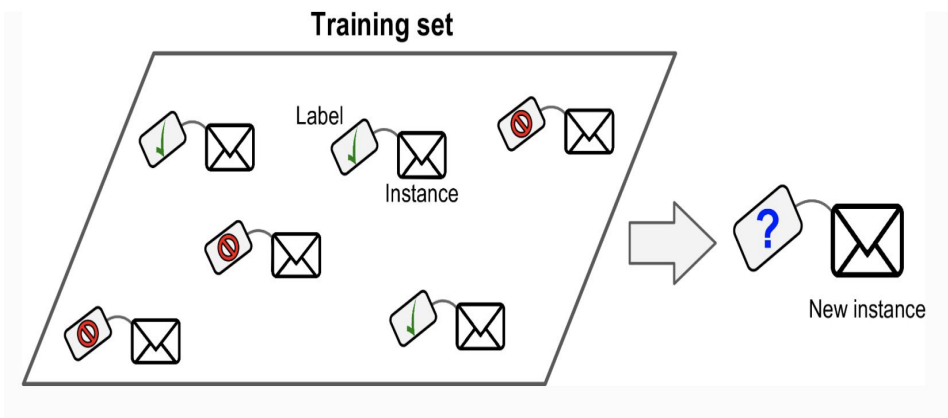


# 1. Aprendizaje Supervisado / No supervisado

- Esta clasificación depende del grado y tipo de supervisión que los algoritmos de ML tienen durante el entrenamiento
- Hay 4 categorías principales: aprendizaje supervisado, no supervisado, semi supervisado y aprendizaje por refuerzo (Reinforcement Learning)

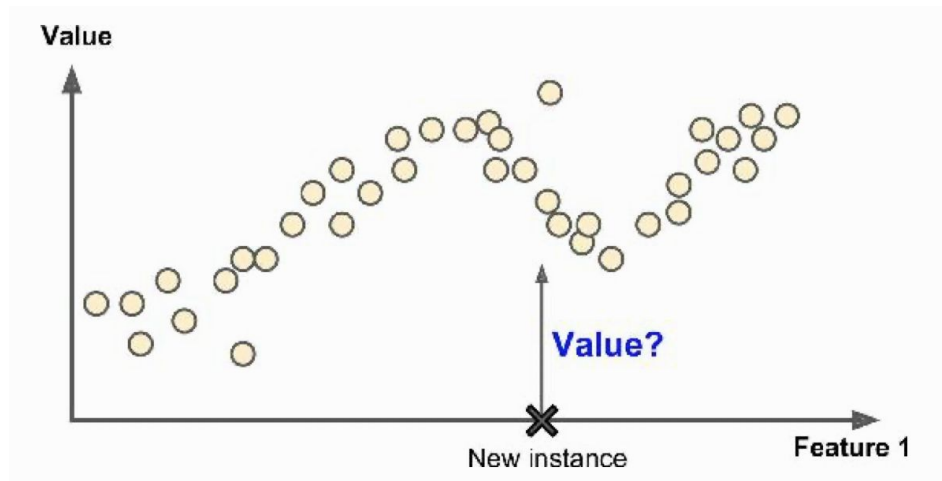
# 1.1 Aprendizaje Supervisado

- Los datos de entrenamiento que se pasan al algoritmo incluyen las soluciones deseadas, llamadas etiquetas
- Una tarea supervisada típica es la **clasificación**
- El filtro de Spam es un ejemplo



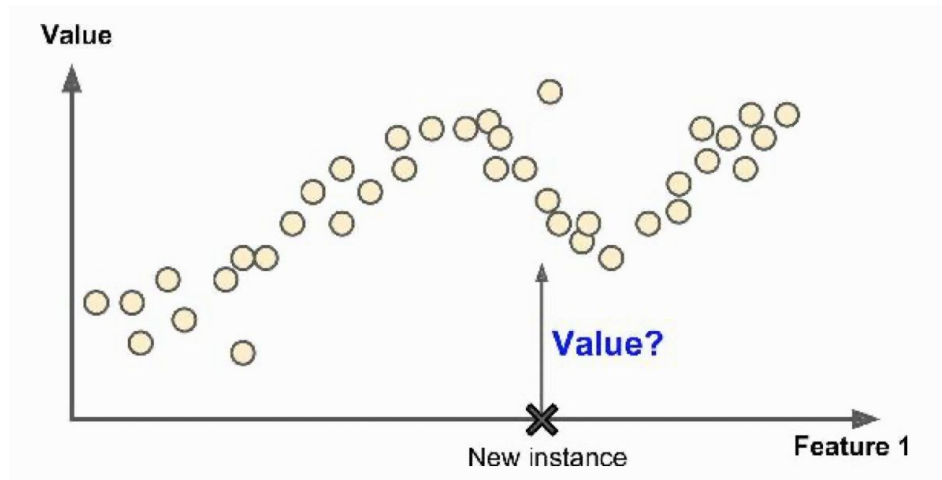
# Aprendizaje Supervisado

- Otra tarea es predecir un valor numérico objetivo, tal como el precio de un carro, dado un conjunto de características (kilometraje, edad, marca, etc.) llamados predictores.
- Esta tarea es llamada de **regresión**
- Para entrenar: muchos ejemplos de carros, incluyendo sus predictores y sus etiquetas (precios)



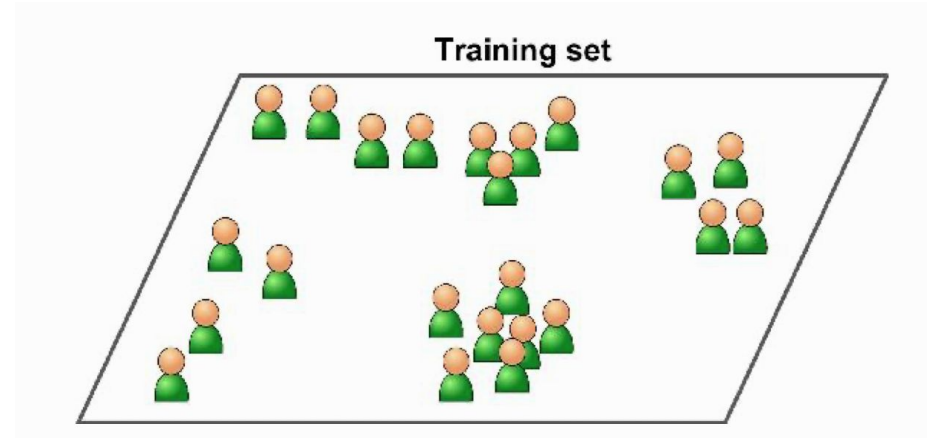
# Algoritmos Supervisados

- K-nearest neighbors
- Regresión Lineal
- Regresión Logística
- Support Vector Machines (SVMs)
- Árboles de decisión y Random Forests
- Redes neuronales
- Redes Neuronales Recurrentes
- Redes Neuronales Convolucionales



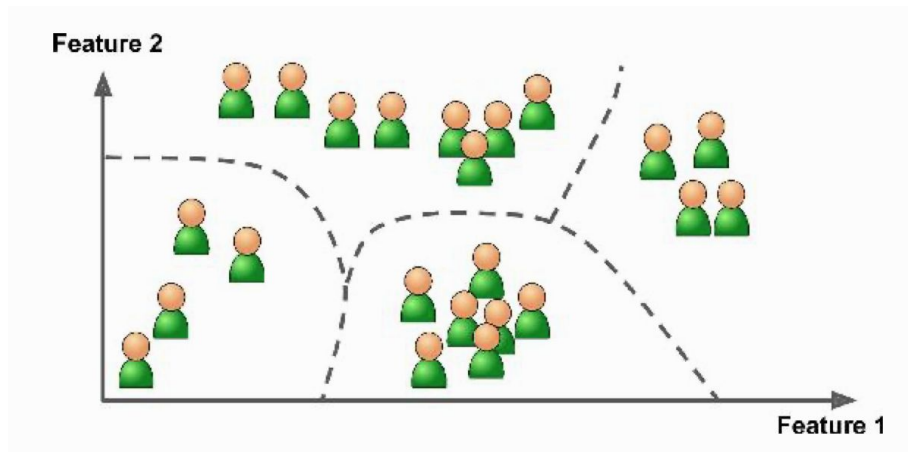
# 1.2 Aprendizaje No Supervisado

- Los datos de entrenamiento no son etiquetados.
- El sistema intenta aprender sin un profesor.
- Típicas tareas no supervisadas:  
Clustering, Visualización y reducción de dimensionalidad, Aprendizaje de reglas de asociación



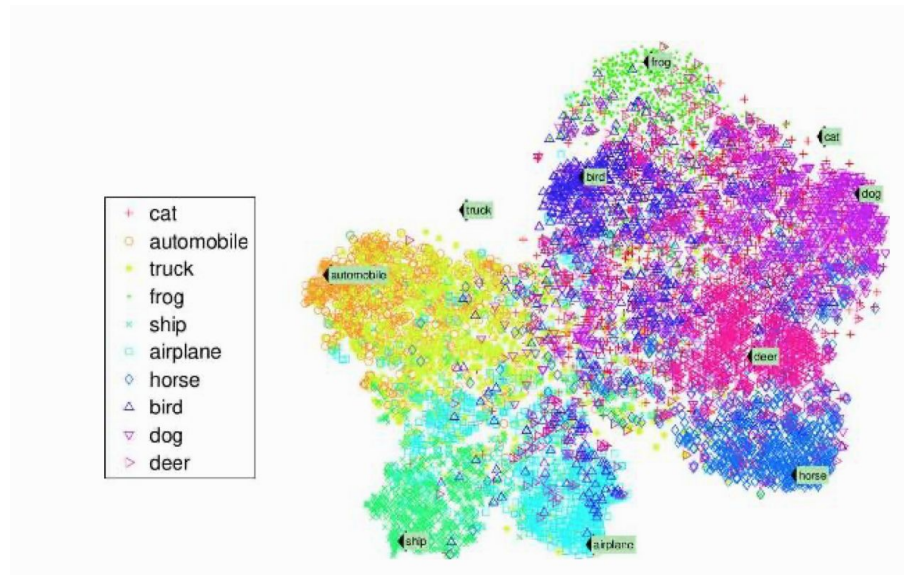
# Aprendizaje No Supervisado: Clustering

- Se tiene datos sobre la visita a un blog
- **Clustering** puede ayudar a detectar grupos similares de visitantes
- El algoritmo encuentra conexiones sin ayuda.
- Podría notar que 40 % de visitantes son varones y leen libros de cómics y generalmente leen el blog en las noches, etc.



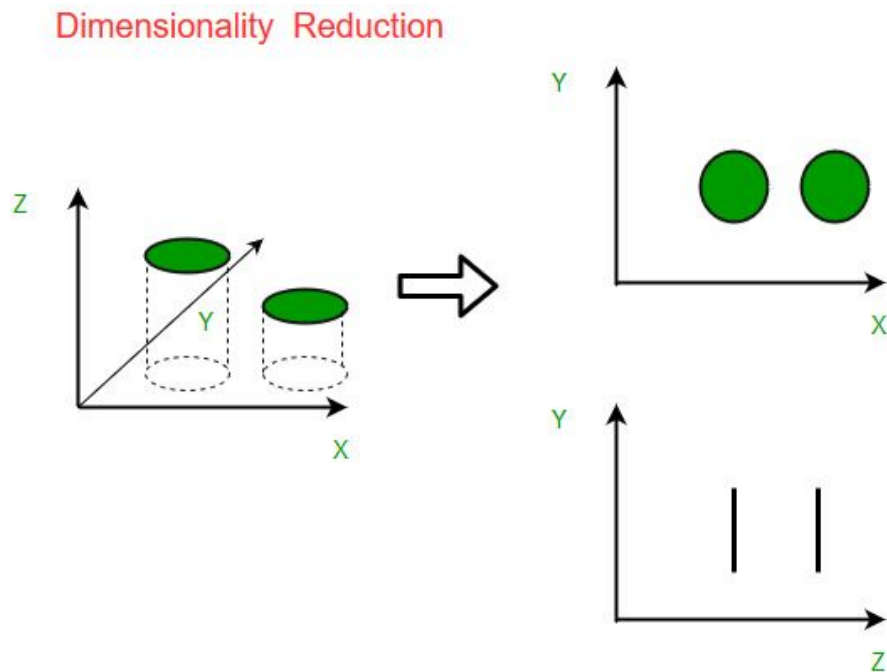
# Aprendizaje No Supervisado: Visualización

- Las entradas son bastantes datos complejos y no etiquetados
- Como resultado se obtiene representaciones en 2D y 3D que pueden ser fácilmente graficadas.
- Estos algoritmos intentan preservar la mayor cantidad de estructura posible
- Se puede entender como los datos son organizados y tal vez identificar patrones inesperados



# Aprendizaje No Supervisado: Reducción Dimensionalidad

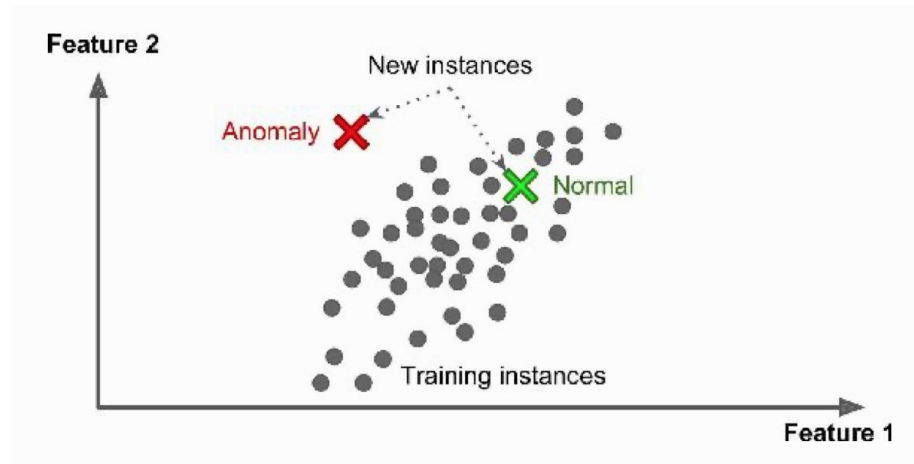
- La meta es simplificar los datos sin perder mucha información
- Una forma es juntar características correlacionadas
- Ejemplo: el kilometraje de un carro, puede estar asociado a su antigüedad. La idea es juntarlos (extracción de características)





# Aprendizaje No Supervisado: Detección de Anomalías

- Para detectar transacciones inusuales en tarjetas de crédito, identificar defectos en manufactura
- El sistema es entrenado con instancias normales y cuando ve una nueva instancia puede indicar si es normal o es una probable anomalía



# Aprendizaje No Supervisado: Aprendizaje de Reglas de Asociación

- La meta es "escarbar" en grandes cantidades de datos y descubrir relaciones interesantes entre atributos (Data mining)
- Ejemplo: ejecutar reglas de asociación en logs de ventas supermercado, puede revelar patrones de compra



# Aprendizaje No Supervisado: Algoritmos

## Clustering:

- K-means
- Hierarchical Cluster Analysis (HCA)
- Expectation Maximization

## Aprendizaje de reglas de asociación:

- Apriori
- Eclat

## Visualización y reducción de dimensionalidad:

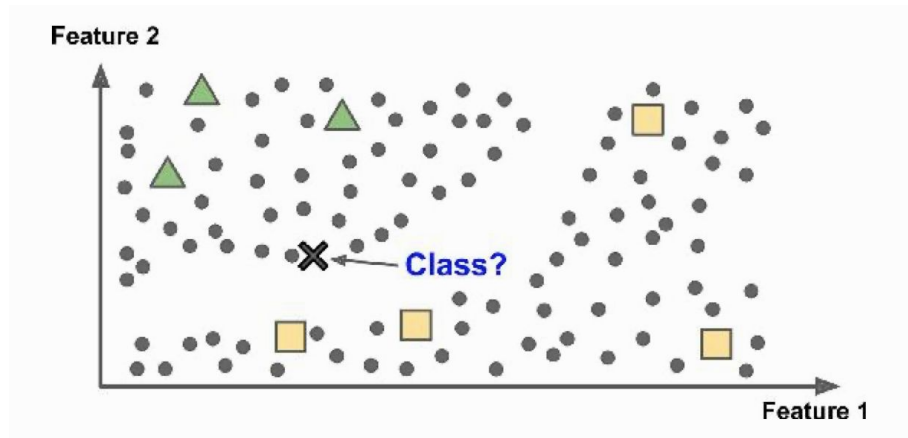
- Principal Component Analysis
- Kernel PCA
- Locally-Linear Embedding (LLE)
- t-distributed Stochastic Neighbor Embedding (t-SNE)

## Detección de anomalías

- One-class SVM
- Isolation Forest

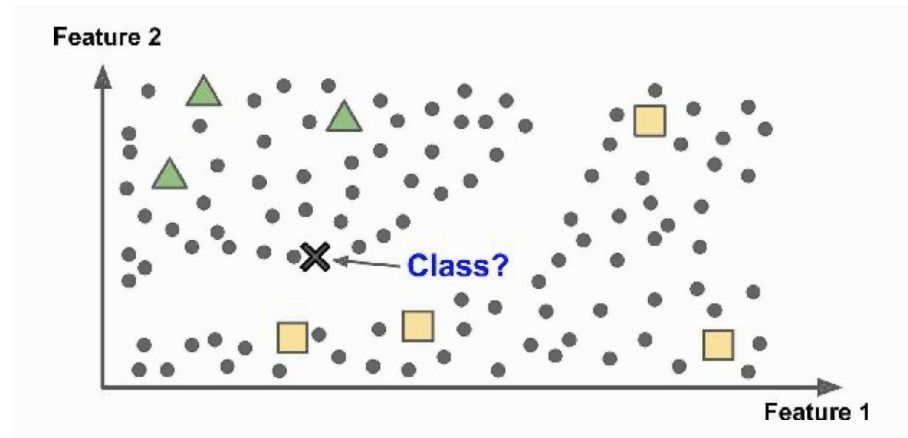
# 1.3 Aprendizaje Semi Supervisado

- Algunos algoritmos lidian con datos de entrenamiento parcialmente etiquetados
- Google fotos: al subir fotos, reconoce que la persona A aparece en fotos 1, 5 y 11 y la persona B aparece en fotos 2, 5 y 7 (parte no supervisada)
- Es necesario indicar quien es la persona



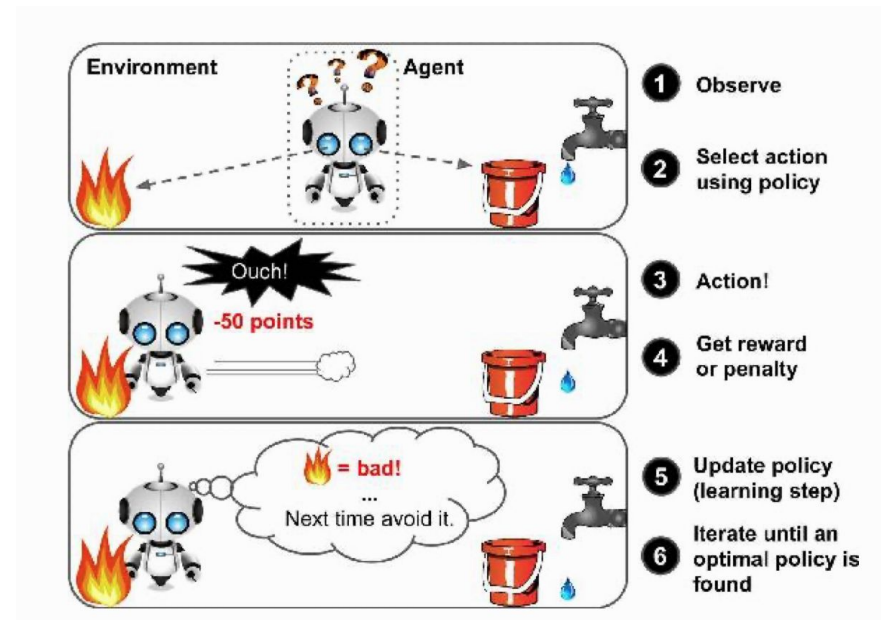
# 1.3 Weak supervision

- Enfoque ML en el que fuentes de alto nivel e incluso con ruido son usadas para crear conjuntos de datos de entrenamiento grandes (IA generativa)
- El objetivo es crear datos rápidamente sin supervisión manual (sin etiquetado manual)



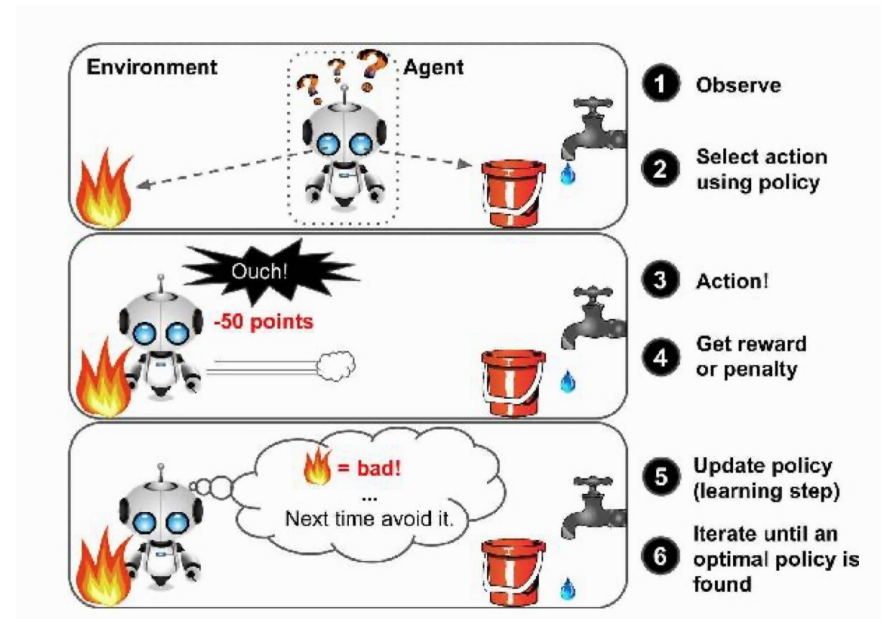
# 1.4 Aprendizaje Por Refuerzo

- Un agente puede observar el ambiente, seleccionar y realizar acciones y obtener recompensas en cambio (o penalidades)
- Debe aprender por sí mismo cuál es la mejor estrategia, política, para obtener la mayor recompensa en el tiempo
- La política define qué acción el agente debe escoger en una situación dada.



# 1.4 Aprendizaje Por Refuerzo

- Muchos robots usan RL para aprender a caminar
- Ejemplo RL: AlphaGo de DeepMind ganó al campeón mundial de Go, Ke Jie en el 2017
- Aprendió su política ganadora al analizar millones de juegos y entonces jugando contra si mismo.

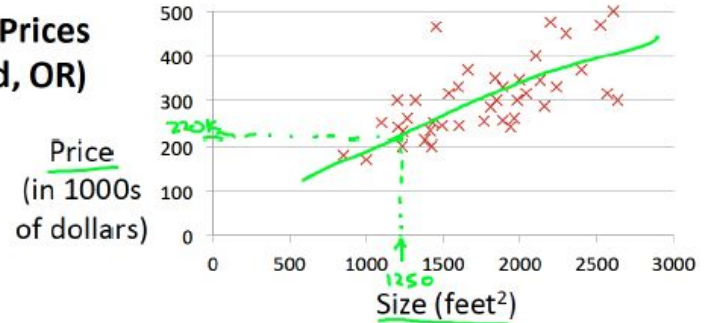


# Regresión

Cuando se tiene la respuestas correcta de cada ejemplo de entrenamiento



Housing Prices  
(Portland, OR)



Regresión: Predice una variable con valores reales



# Clasificación

## Clasificación:

- Emails: Spam o no Spam
- Transacción bancaria: fraudulenta o no
- Tumor: maligno o no

$y \in \{0, 1\}$  0: clase negativa, 1: clase positiva

$y \in \{0, 1, 2, 3, 4\}$

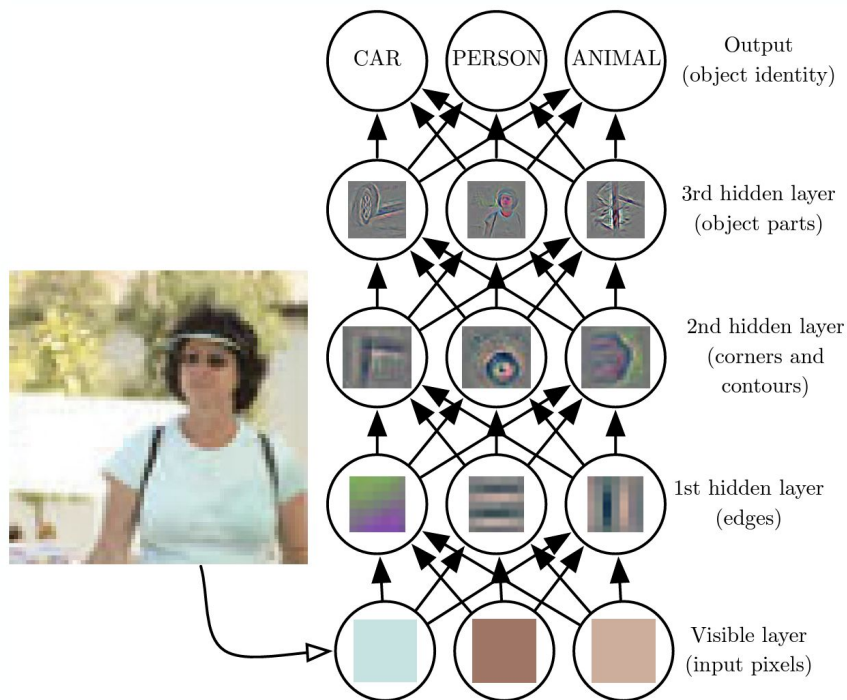
# Definición Deep Learning

Son métodos de aprendizaje con múltiples niveles de representación, obtenidos al componer módulos simples no lineales, cada uno de los cuales transforma la representación de un nivel (comenzando con la entrada sin procesar o raw) a una representación o abstracción de nivel superior.

# Arquitectura Deep Learning

- Con múltiples capas no lineales, digamos una profundidad de 5 a 20, un sistema puede implementar funciones extremadamente complejas que son sensibles a detalles minuciosos y
- Insensibles a variaciones como el fondo, la pose, la iluminación y los objetos circundantes.

# Profundidad: composición repetida



# Proceso de "Aprendizaje"

