Machine Learning AI3002

Lecture 01

What is Machine learning in simple words?

- Machine learning is a branch of artificial intelligence (AI) and computer science that focuses on the use of data and algorithms to imitate the way that humans learn.

What is Machine learning in simple words?

- Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

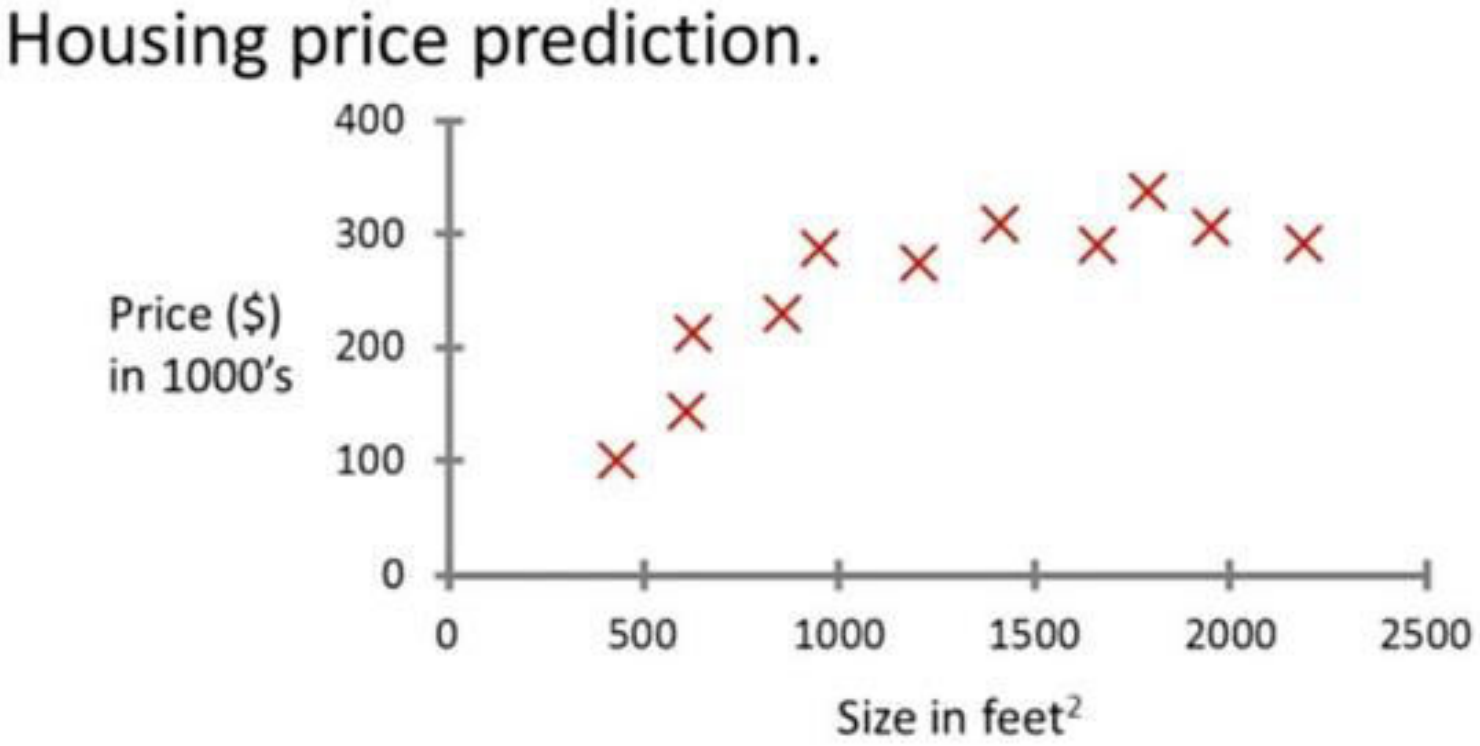
What is Machine learning in simple words?

- Tom Mitchell (1998) A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.
- Suppose your email program watches which email you do or do not mark as spam, and based on that learns how to better filter spam.
  - Classifying emails as spam or not spam. (T)
  - Watching you label emails as spam or not spam. (E)
  - The number (or fraction) of emails correctly classified as spam/not spam. (P)

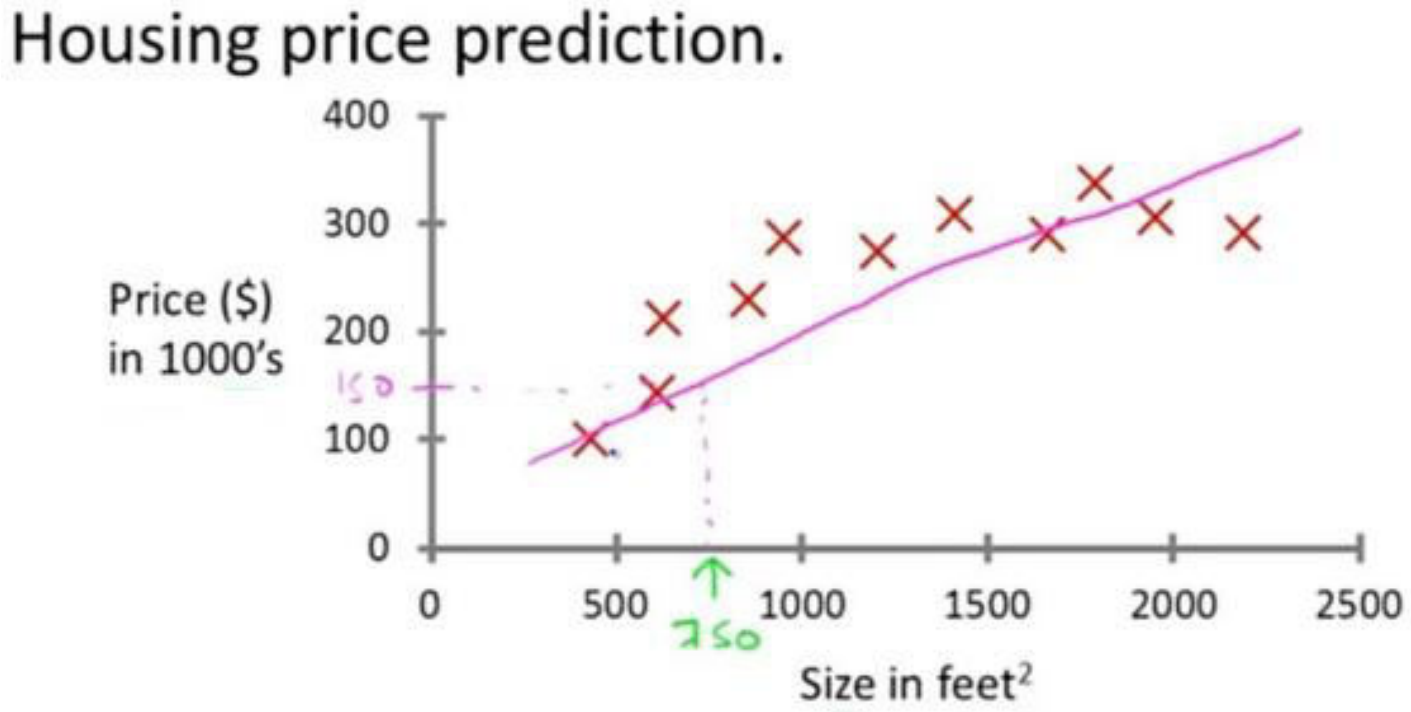
Types of Machine Learning Algorithms

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

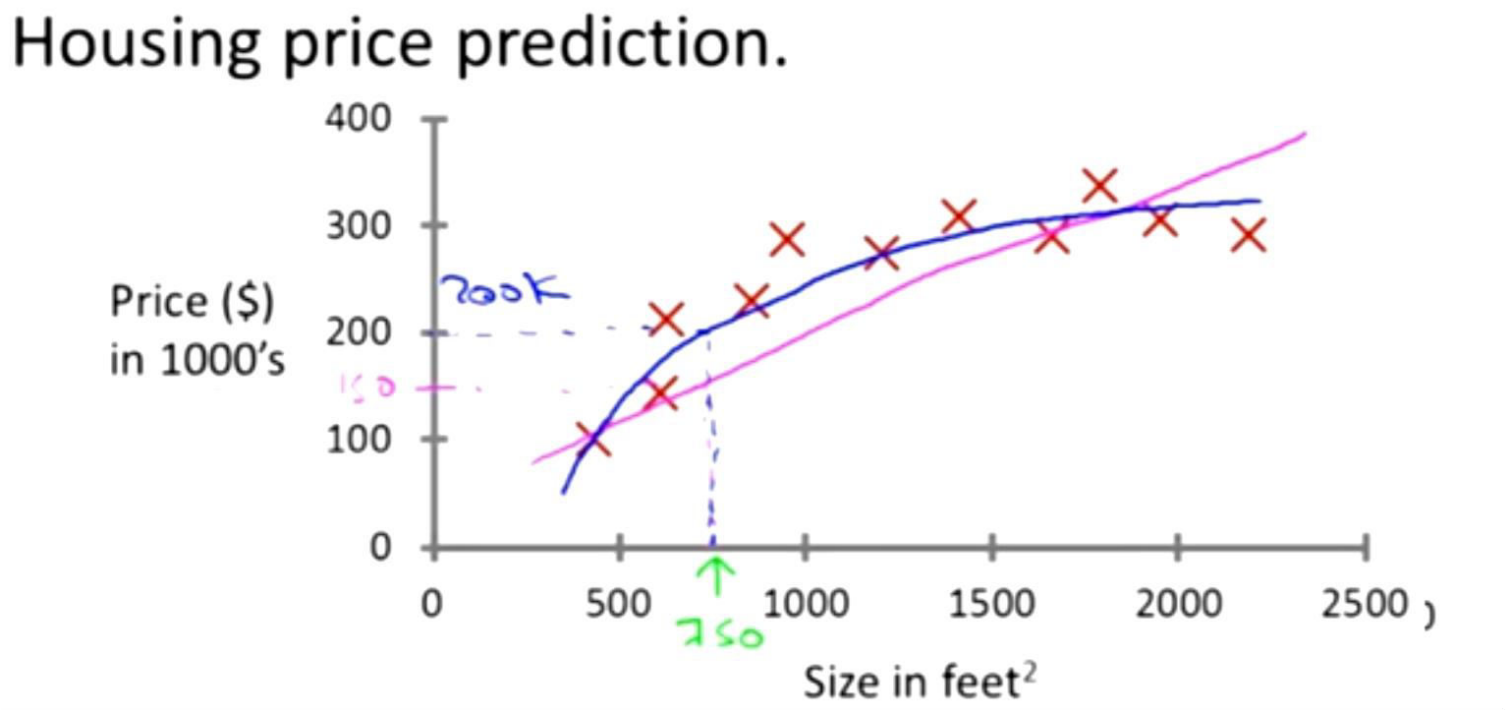
Supervised Learning (Regression)



Supervised Learning (Regression)



Supervised Learning (Regression)



Regression: Predict continuous valued output (price)

Supervised Learning (Classification)

- An emergency room in a hospital measures 17 variables (e.g., blood pressure, age, etc.) of newly admitted patients. A decision must be taken whether to put the patient in an intensive-care unit. Due to the high cost of ICU, those patients who may survive less than a month are given higher priority. The problem is to predict high-risk patients and discriminate them from low-risk patients.

Supervised Learning (Classification)

- A credit card company typically receives thousands of applications for new cards. The application contains information regarding several different attributes, such as annual salary, any outstanding debts, age etc. The problem is to categorize applications into those who have good credit, bad credit, or fall into a gray area (thus requiring further human analysis).

Supervised Learning (Classification)

- Data: It has k attributes A1 , … Ak . Each tuple (case or example) is described by values of the attributes and a class label.
- Goal: To learn rules or to build a model that can be used to predict the classes of new (or future or test) cases.
- The data used for building the model is called the training data.

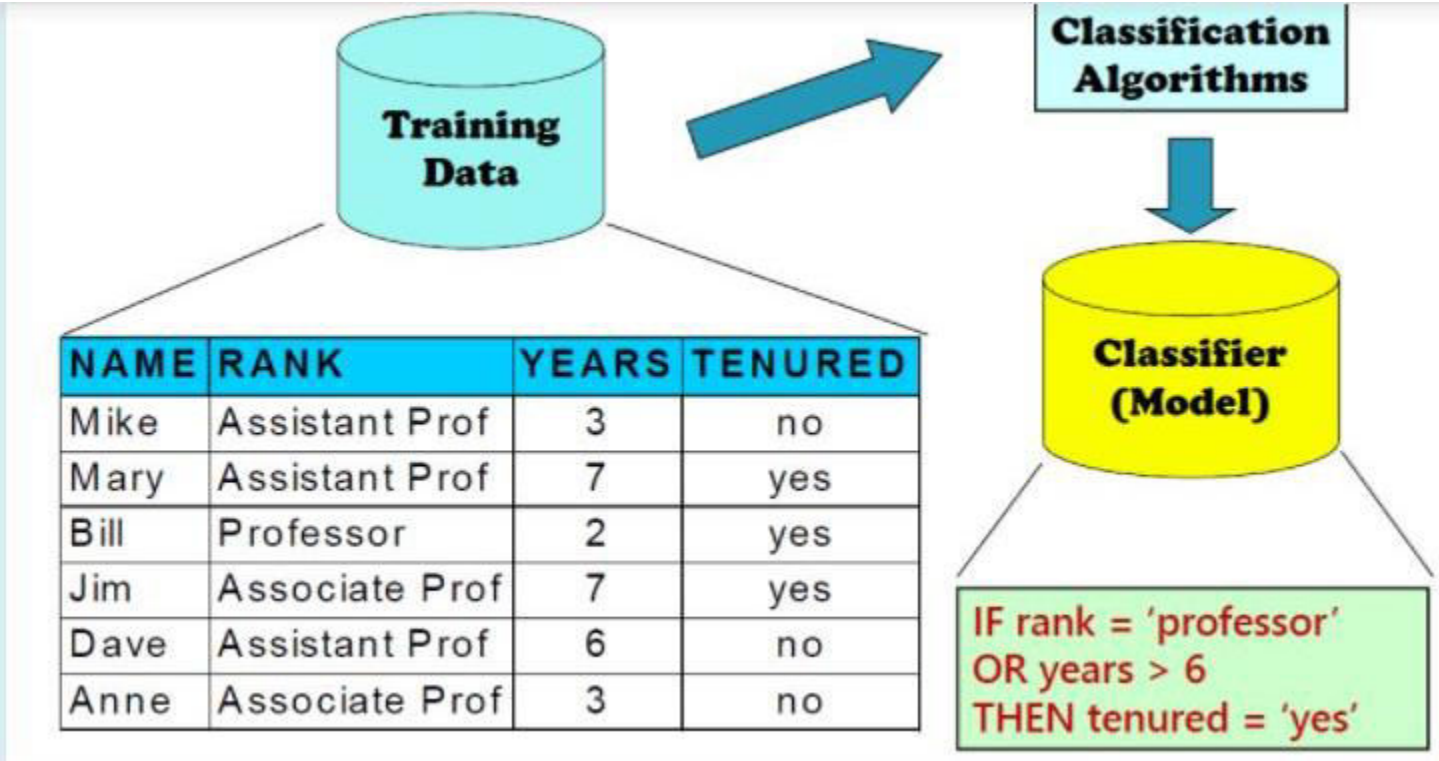
Supervised Learning (Classification)

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Classification Two-Step Process

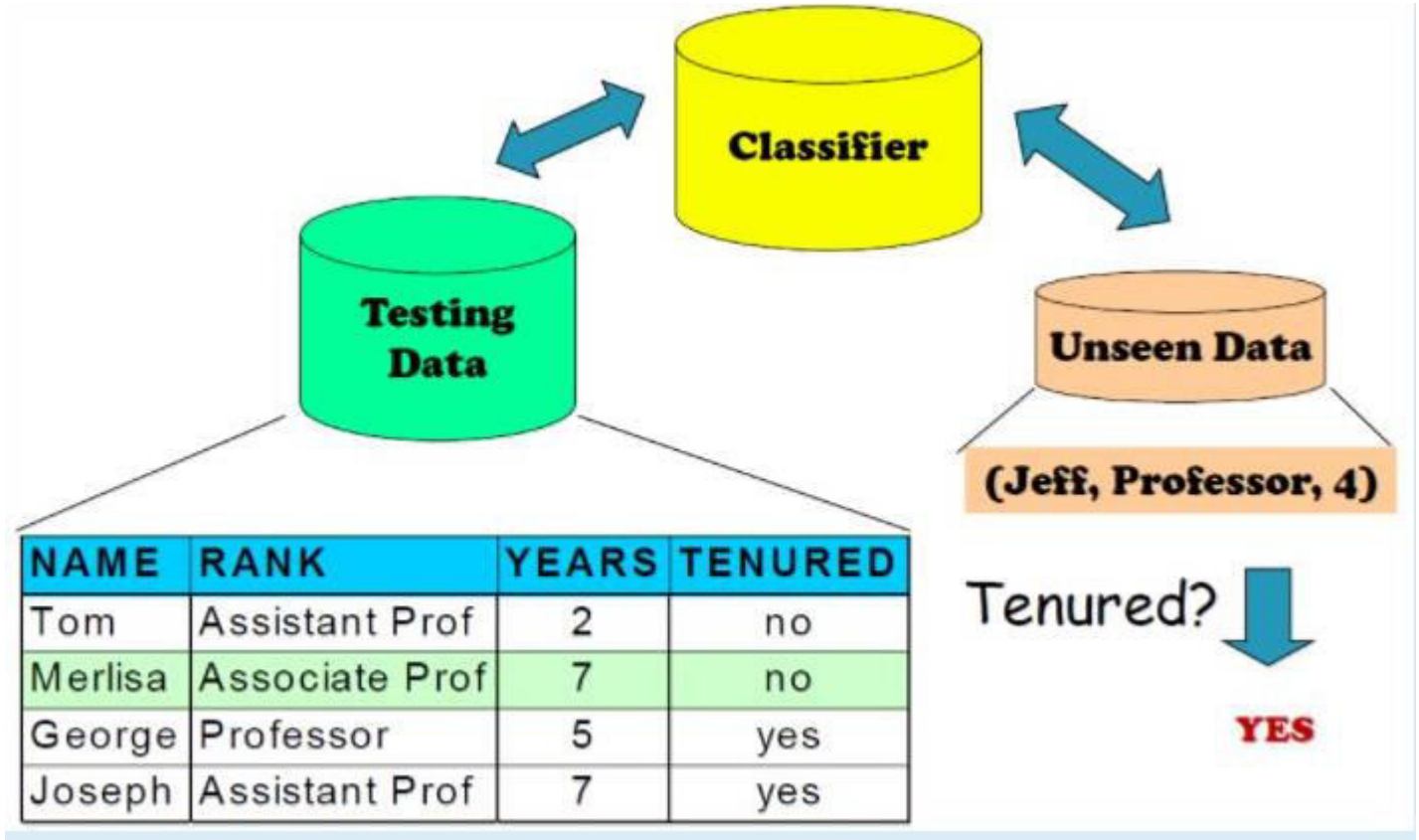
- Model construction: describing a set of predetermined classes based on a training set. It is also called learning.
  - Each tuple/sample is assumed to belong to a predefined class
  - The model is represented as classification rules, decision trees, or mathematical formulae .
- Model usage: for classifying future test data/objects
  - Estimate accuracy of the model
    - The known label of test example is compared with the classified result from the model
    - Accuracy rate is the % of test cases that are correctly classified by the model
  - If the accuracy is acceptable, use the model to classify data tuples whose class labels are not known.

Classification Process (1): Model Construction



NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no

Classification Process (2): Use the Model in Prediction



NAME	RANK	YEARS	TENURED
Tom	Assistant Prof	2	no
Merlisa	Associate Prof	7	no
George	Professor	5	yes
Joseph	Assistant Prof	7	yes

Supervised vs Unsupervised Learning

- Supervised learning: classification is seen as supervised learning from examples.
  - Supervision: The training data (observations, measurements, etc.) are accompanied by labels indicating the classes of the observations/cases.
  - New data is classified based on the training set
- Unsupervised learning (clustering)
  - The class labels of training data is unknown
  - Given a set of measurements, observations, etc. with the aim of establishing the existence of classes or clusters in the data

Clustering

- Clustering is a technique for finding similar groups in data, called clusters. i.e.,
  - It groups data instances that are similar to (near) each other in one cluster and data instances that are very different (far away) from each other into different clusters.
- Clustering is often called an unsupervised learning task as no class values denoting an a priori grouping of the data instances are given, which is the case in supervised learning.

What is clustering for?

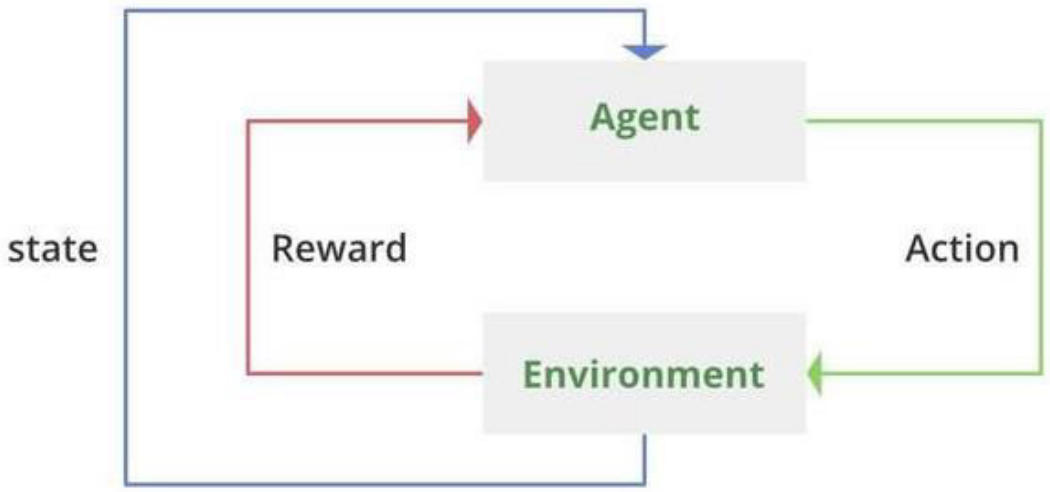
Let us see some real-life examples

- Example 1: groups people of similar sizes together to make “small”, “medium” and “large” T-Shirts.
  - Tailor-made for each person: too expensive
- Example 2: In marketing, segment customers according to their similarities
  - To do targeted marketing.

What is clustering for?

- Example 3: Given a collection of text documents, we want to organize them according to their content similarities,
  - To produce a topic hierarchy
  - It has a long history, and used in almost every field, e.g., medicine, psychology, botany, sociology, biology, archeology, marketing, insurance, libraries, etc.
  - In recent years, due to the rapid increase of online documents, text clustering becomes important.

Reinforcement learning



- In reinforcement learning, an agent interacts with an environment with an objective to maximize its total reward.
- The agent takes an action based on the environment state and the environment returns the reward and the next state. The agent learns from trial and error, initially taking random actions and over time identifying the actions that lead to long-term rewards.