

# MACHINE LEARNING

Esempi di intelligenza artificiale fatti inizialmente:

- Ricerca google -> raccomandazione
- Apple foto che raggruppa le foto in base alle persone -> Clusterizzazione di immagini
- Blocco spam -> classificazione

**Il Machine learning** è una parte dell'intelligenza artificiale con l'obiettivo di creare delle macchine "intelligenti"

- Calcolatori capaci di risolvere dei problemi che noi avremmo capacità a identificare
  - Examples:
    - ♦ Applications that can't be programmed by hand
      - E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision
    - ♦ Database mining
      - Large datasets from growth of automation/web
      - E.g., web click data, medical records, biology, engineering
    - ♦ Self-customizing programs
      - E.g., Amazon, Netflix product recommendations
    - ♦ ...

**Machine learning definition:**

- *"Campo di studio che dà l'abilità ai computer di imparare senza essere esplicitamente programmati"* (Arthur Samuel)
  - *Insegnò ad una macchina a giocare a scacchi facendo giocare la macchina contro la stessa macchina*
- *"Un programma impara da un'esperienza  $E$  rispetto ad un certo task  $T$  dando una qualche misura di performance  $P$ , se la sua performance su  $T$ , come misurato da  $P$ , migliora con l'esperienza  $E$ "* (Tom Mitchell)

**Machine learning algorithms**

- Supervised learning
- Unsupervised learning
- Others

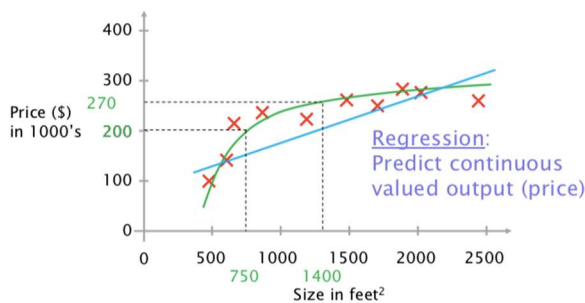
## Supervised learning

Il termine supervised si riferisce al fatto che noi diamo all'algoritmo un dataset formale, che abbia una serie di passi che porteranno a minimizzare (ottimizzare) una funzione di costo.

Dunque, insegno al modello a dare delle risposte, fornendogli degli esempi di cui forniamo le cosiddette risposte corrette.

Il compito dell'algoritmo è quello di fornire risposte ulteriori a domande diverse sullo stesso problema. Questo tipo di **problemi** si chiamano anche di **regressione** perché cerchiamo di predire un valore di output continuo (non discreto) dati gli input.

### ▪ Housing price prediction

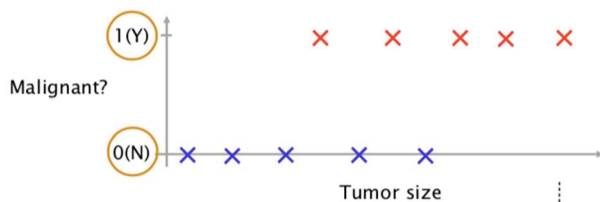


Partendo dai valori che avviamo, vogliamo:

- Calcolare una retta che fitti al meglio i dati che abbiamo -dataset- (regressione lineare)
- La curva fitta meglio i dati e quindi vorrei trovare la curva per poter predire il prezzo a cui posso vendere la casa in base ai piedi quadrati.

Un altro tipo di problema è quello di **classificazione**: discreto.

### ▪ Breast cancel (malignant, benign)

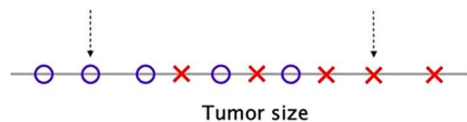


Esempi:

- Capire se tumore benigno o maligno
- Capire se giornata piovosa/soleggiata/nuvolosa

Classification:  
Discrete valued output (in this case)

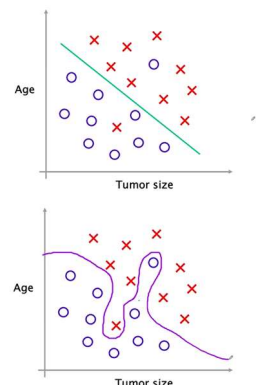
Not necessarily just two classes (e.g., 0, 1, 3, 4, etc.)



Si può anche rappresentare su un solo asse

All'aumentare del nr di aspetti da considerare, si complica lo schema e avremo:

- Decision boundary: linea che mi permette di dire se benigno o maligno
  - o Retta: potrebbe essere grossolana
  - o Curva che fitta perfettamente: algoritmo non capace di generalizzare, fitta benissimo quei dati ma se lo si sottopone ad altri dati, probabilmente sbaglia.
    - Algoritmo deve adattarsi ai dati ma non troppo.



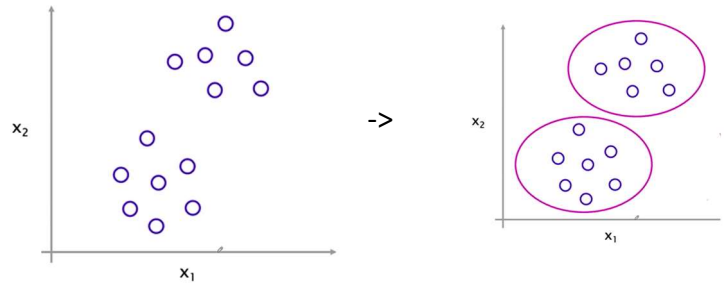
## Unsupervised learning

Il dataset non è etichettato e si chiede all'algoritmo di individuare una qualche struttura e/o suddivisione.

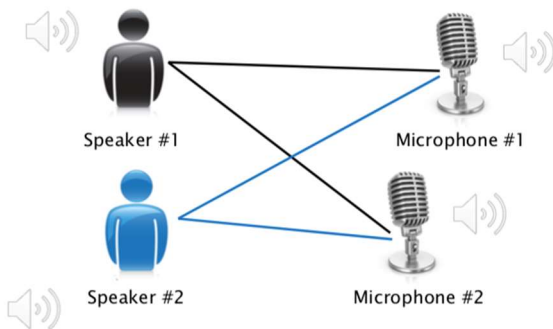
Ad esempio se l'algoritmo raggruppa i dati in insiemi si ha un algoritmo di clustering.

Esempi di utilizzi sono:

- Google News: articoli simili l'uno all'altro vengono raggruppati
- Genomica
- Reti: cluster di calcolatori creati mettendo vicini macchine di utenti simili o vicini tra loro per minimizzare lo scambio di informazioni
- Social network: utilizzano tecniche di clustering per avvicinarsi a qualcun altro nella rete di amicizie.
- Segmentazione del mercato
- Analisi dei dati astronomici



### ▪ Cocktail party problem



Esempio delle due persone che contano da 1 a 10.