



**unimc**  
UNIVERSITÀ DI MACERATA



Week 2 – Lezione 3

# Introduzione al machine learning e applicazioni in finanza

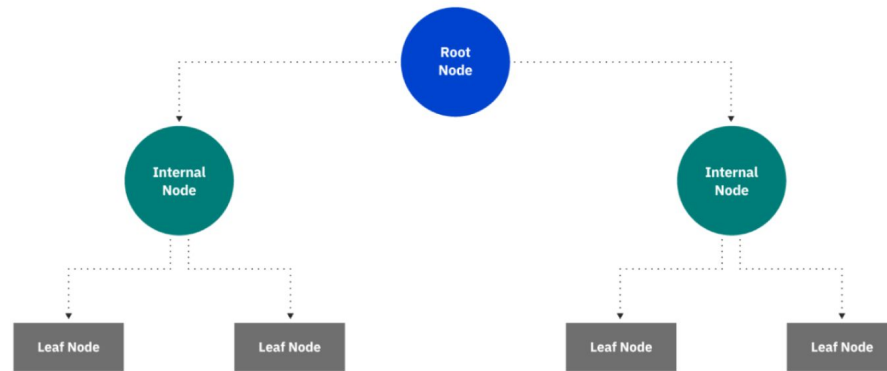
## Classificazione/Regressione

Prof. Luca Romeo

# Modelli di classificazione e/o regressione

## Albero decisionale

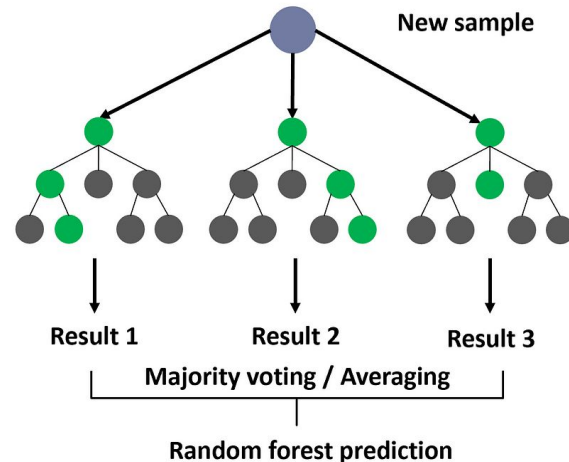
L'albero decisionale è un modello di machine learning utilizzato per task di classificazione e regressione che suddivide la classificazione in una serie di scelte gerarchiche per ogni voce del vettore delle caratteristiche



# Modelli di classificazione e/o regressione

## Random forest

La Random forest rientra nella categoria degli algoritmi di "**ensemble learning**", che combinano diversi modelli per migliorare le prestazioni complessive. In questo caso l'ensemble è costituito da un **insieme di alberi decisionali**.



# Modelli di classificazione e/o regressione

## Support Vector Machine

Una Support Vector Machine (**SVM**) è un modello machine learning utilizzato per **entrambi** problemi di classificazione e regressione.

Nei task di classificazione, anche se le SVM possono essere estese anche a problemi multi-classe, solitamente sono utilizzate per task binari.

In generale, una SVM ha l'**obiettivo** di trovare un **iperpiano ottimale** (una superficie di decisione) che separa i dati di addestramento in classi diverse (i.e., l'iperpiano di separazione ottimale).

# Modelli di classificazione e/o regressione

## Support Vector Machine – L'iperpiano

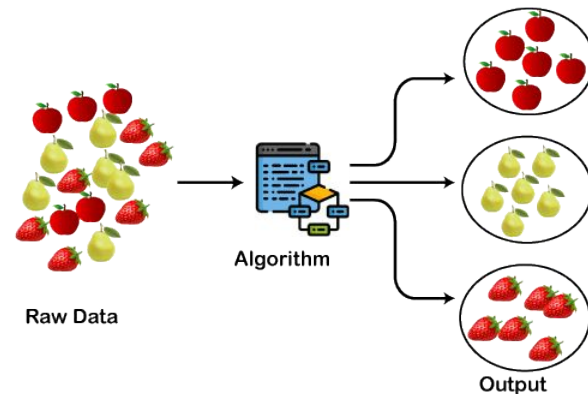
Generalizzazione del concetto di piano o retta in spazi di dimensioni superiori.

- In uno spazio **2D**, è una **retta**.
- In uno spazio **3D**, è un **piano**.
- In uno spazio **n-dimensionale**, è una superficie di dimensione  $n-1$  che divide lo spazio in due parti.
- Per una SVM, l'iperpiano è la superficie che separa i dati delle due classi.

# Machine Learning non supervisionato

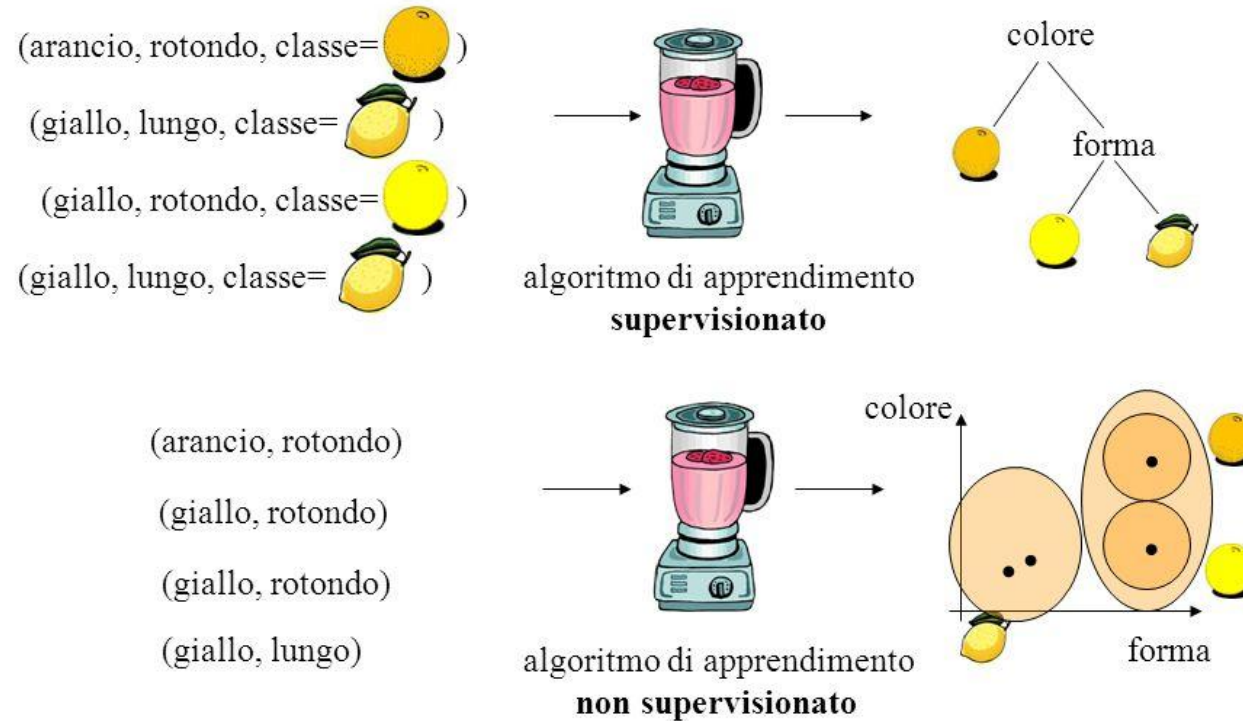
In questo metodo **non è necessario supervisionare** o dare in input al modello dei dati etichettati. L'algoritmo inizia ad imparare dai dati senza una guida o un controllo a priori.

Le macchine con algoritmi di apprendimento non supervisionati scoprono schemi per trovare risultati utili.



L'algoritmo non può etichettare gli oggetti ma può raggrupparli in gruppi simili

# Un confronto visivo



# Come scegliere il giusto approccio?

Ogni workflow basato su machine learning comincia con 3 quesiti principali:

- Con che tipo di dati lavoriamo?
- Cosa vogliamo ottenere dai dati?
- In che ambito lavoriamo?

# Come scegliere il giusto approccio?

Scegliamo **supervisionato...**



Se abbiamo bisogno di addestrare un modello per fare una **previsione** o **classificazione**.

Scegliamo **non supervisionato...**



Quando vogliamo capire la **distribuzione** dei dati, il **raggruppamento** dei dati sulla base di caratteristiche simili o la riduzione delle loro dimensioni, al fine di fornirne una versione più sintetica ed efficace.