



unimc
UNIVERSITÀ DI MACERATA



Week 2 – Lezione 2

Introduzione al machine learning e applicazioni in finanza

Classificazione

Prof. Luca Romeo

Regressione e classificazione

Gli algoritmi di apprendimento supervisionato possono gestire due tipi principali di output:

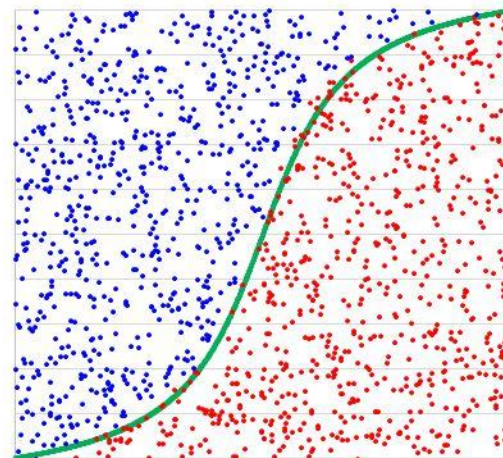
- **Classificazione:** Utilizzata quando l'output desiderato è discreto, ovvero appartiene a categorie o classi predefinite. Ad esempio, classificare un'email come "spam" o "non spam".
- **Regressione:** Utilizzata quando l'output desiderato è continuo. Ad esempio, prevedere il prezzo di una casa o la temperatura.

Modelli di classificazione

Regressione Logistica

La Regressione Logistica (LR) è un modello lineare impiegato per task di **classificazione binaria**. Il suo scopo è **predire la probabilità** che un campione appartenga a una delle due classi.

- La funzione logistica (o sigmoide) prende l'output di una funzione lineare (un numero reale) e lo trasforma in una probabilità p compresa tra 0 e 1.
- Se $p \geq 0.5$, il campione viene assegnato alla classe positiva; altrimenti, viene assegnato alla classe negativa.



Modelli di classificazione

K-Nearest Neighbors (KNN)

Il K-Nearest Neighbors (KNN) è un algoritmo di classificazione supervisionato basato sulla **vicinanza** (distanza) dei dati.

Se due campioni presentano caratteristiche simili, è probabile che appartengano alla stessa classe. In altre parole, oggetti simili tendono ad essere vicini nello spazio delle caratteristiche.

