

10. Introduzione al machine learning

Corso di Python per il Calcolo Scientifico

Outline

- Introduzione al machine learning
- Sistemi ad apprendimento supervisionato
- Sistemi ad apprendimento non supervisionato
- Sistemi di reinforcement learning

Introduzione al machine learning

- Il machine learning ci offre un modo **più efficace** di risolvere problemi complessi
- Prevede un processo di **addestramento** di un software, chiamato **modello**, che viene utilizzato per fare predizioni a partire da un insieme di dati
- È contrapposto all'approccio tradizionale. Ad esempio:
 - un **sistema di previsioni meteo tradizionale** si affida a **complesse equazioni** che stimano valori di umidità, pressione e pioggia, mentre...
 - un **sistema di previsioni meteo basato su machine learning** si affida ad un **algoritmo** in grado di fare previsioni a partire da un addestramento su grosse quantità di dati.
- Esistono sistemi ad **apprendimento supervisionato, non supervisionato** e di **reinforcement learning**.

Sistemi ad apprendimento supervisionato

- I sistemi ad **apprendimento supervisionato** partono da un insieme di dati **etichettati**
- In pratica, un esperto di dominio definisce il valore da predire a partire per un certo campione, ed il modello è delegato ad individuare le relazioni che intercorrono tra le feature in input ed il valore atteso

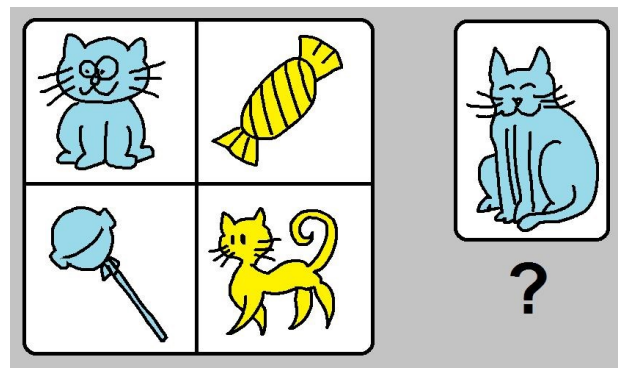
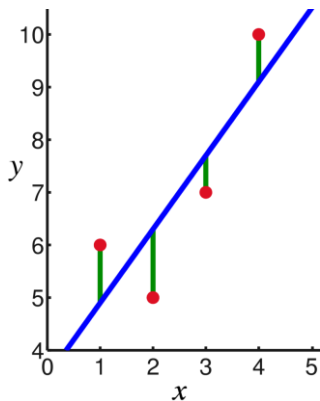
Feature ↓

Label ↓

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500

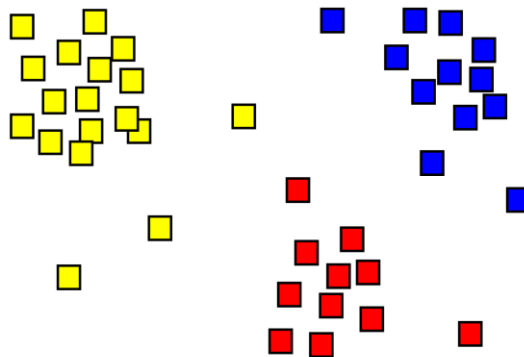
Sistemi ad apprendimento supervisionato

- I sistemi ad apprendimento supervisionato si dividono in due categorie principali
- I **modelli di regressione** predicono un dato di tipo numerico
- I **modelli di classificazione** predicono un dato di tipo categorico (ovvero una **classe**)

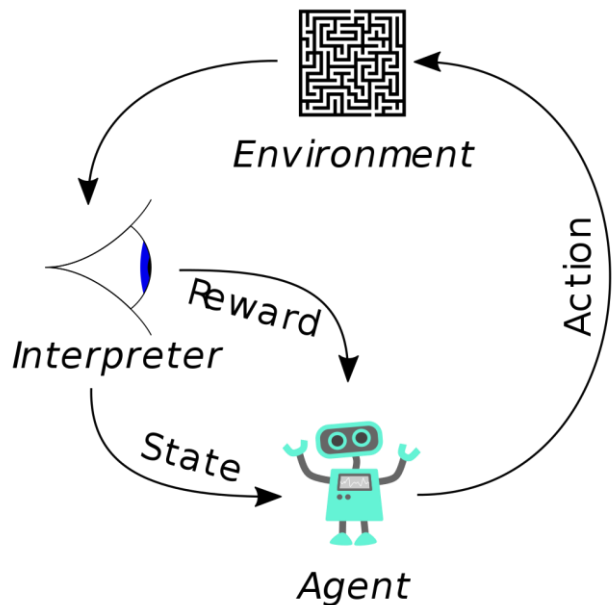


Sistemi ad apprendimento non supervisionato

- I sistemi ad **apprendimento non supervisionato** inferiscono la suddivisione dei dati senza alcun apporto esterno
 - Non necessitano quindi di una etichettatura effettuata da esperti di dominio
- Tipici esempi sono gli algoritmi di **clustering**
 - Questi algoritmi suddividono lo spazio dei campioni sulla base di **metriche di distanza** o **criteri di agglomerazione**



Sistemi di reinforcement learning



- I sistemi di **reinforcement learning** sono concettualmente differenti da quelli di apprendimento
- Prevedono che un agente effettui una **azione** all'interno di un **ambiente** sulla base di una **ricompensa**
- La ricompensa viene valutata sulla base dello stato attuale dell'ambiente

Domande?

42