

Machine Learning

UniShare

Davide Cozzi
@dlcgold

Indice

1	Introduzione	2
2	Introduzione al ML	3
2.1	Concept learning	4

Capitolo 1

Introduzione

Questi appunti sono presi a lezione. Per quanto sia stata fatta una revisione è altamente probabile (praticamente certo) che possano contenere errori, sia di stampa che di vero e proprio contenuto. Per eventuali proposte di correzione effettuare una pull request. Link: <https://github.com/dlccgold/Appunti>.

Capitolo 2

Introduzione al ML

Il **Machine Learning** (*ML*) è sempre più diffuso nonostante sia nato diversi anni fa.

Un **sistema di apprendimento automatico** ricava da un *dataset* una conoscenza non fornita a priori, descrivendo dati non forniti in precedenza. Si estrapolano informazioni facendo assunzioni sulle informazioni sistema già conosciute, creando una **classe delle ipotesi H**. Si cercano ipotesi coerenti per guidare il sistema di apprendimento automatico. Bisogna però mettere in conto anche eventuali errori, cercando di capire se esiste davvero un'ipotesi coerente e, in caso di assenza, si cerca di approssimare. In quest'ottica bisogna mediare tra **fit** e **complessità**. Ogni sistema dovrà cercare di mediare tra questi due aspetti, un *fit* migliore comporta alta *complessità*. Si ha sempre il rischio di **overfitting**, cercando una precisione dei dati che magari non esiste. Si ha un **generatore di dati** ma il sistema non ha conoscenza della totalità degli stessi.

Durante l'**apprendimento** si estrapolano dati da **istanze di addestramento o test**. Quindi:

- si ricevono i dati di addestramento
- il sistema impara ad estrapolare partendo da quei dati
- si ricevono dati di test su cui si estrapola

L'ipotesi da apprendere viene chiamata **concetto target** (tra tutte le ipotesi possibili identifico quella giusta dai dati di addestramento).

Si hanno tre tipi di apprendimento:

1. **apprendimento supervisionato**, con un *insegnante* che comunica l'output corretto

2. **apprendimento non supervisionato**, dove si riconosce *schemi* nell'input senza indicazioni sui valori in uscita. Questo tipo si basa sui *cluster*. Non c'è target
3. **apprendimento per rinforzo**, dove bisogna apprendere sulla base della risposta dell'ambiente alle proprie azioni. Si lavora in *modo continuo*, aggiornando le ipotesi con l'arrivo dei dati

2.1 Concept learning

In questo settore il dato è un vettore booleano e il dato che viene richiesto in output è anch'esso un booleano. Il target dell'addestramento