
ANNO ACCADEMICO 2024/2025

Apprendimento Automatico

Teoria

Altair's Notes



UNIVERSITÀ
DI TORINO



DIPARTIMENTO DI INFORMATICA

CAPITOLO 1

INTRODUZIONE

PAGINA 5

- 1.1 Le basi del machine learning
Tasks — 7 • Modelli — 8 • Features — 9

5

CAPITOLO 2

TEST2

PAGINA 11

Premessa

Licenza

Questi appunti sono rilasciati sotto licenza Creative Commons Attribuzione 4.0 Internazionale (per maggiori informazioni consultare il link: <https://creativecommons.org/version4/>).



Formato utilizzato

Box di "Concetto sbagliato":

Concetto sbagliato 0.1: Testo del concetto sbagliato

Testo contenente il concetto giusto.

Box di "Corollario":

Corollario 0.0.1 Nome del corollario

Testo del corollario. Per corollario si intende una definizione minore, legata a un'altra definizione.

Box di "Definizione":

Definizione 0.0.1: Nome delle definizioni

Testo della definizione.

Box di "Domanda":

Domanda 0.1

Testo della domanda. Le domande sono spesso utilizzate per far riflettere sulle definizioni o sui concetti.

Box di "Esempio":

Esempio 0.0.1 (Nome dell'esempio)

Testo dell'esempio. Gli esempi sono tratti dalle slides del corso.

Box di "Note":

Note:-

Testo della nota. Le note sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive.

Box di "Osservazioni":

Osservazioni 0.0.1

Testo delle osservazioni. Le osservazioni sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive. A differenza delle note le osservazioni sono più specifiche.

1

Introduzione

1.1 Le basi del machine learning

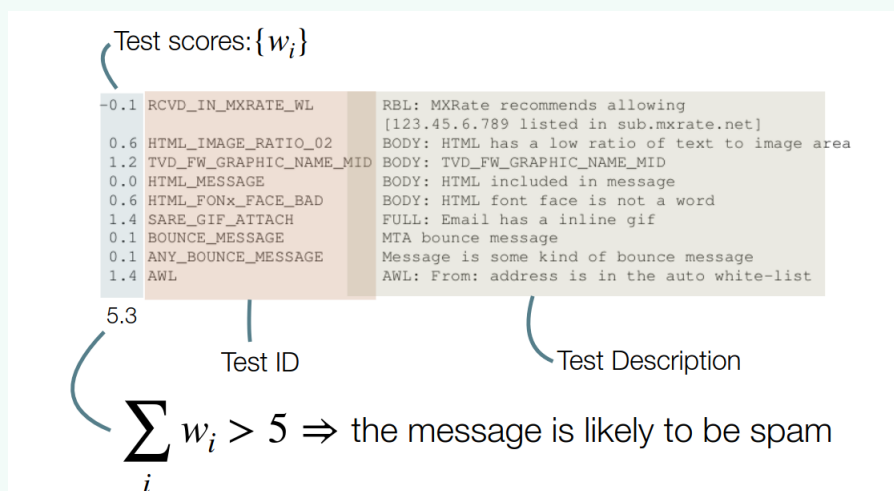
Gli ingredienti del machine learning:

- ⇒ *Task*: specifica di cosa si vuole fare;
- ⇒ *Modelli*: il modello matematico per affrontare un determinato task;
- ⇒ *Features*: il modo con cui sono descritti gli esempi.

Note:-

L'*apprendimento automatico* ruota attorno all'idea di estrarre una regola generale per risolvere un problema a partire da problemi già risolti.

Esempio 1.1.1 (Etichettatura delle email spam)



SpamAssassin è un filtro open-source usato per filtrare lo spam. Esso non lavora sul testo, ma su alcune *feature* della mail.

E-mail	x_1	x_2	Spam?	$4x_1 + 4x_2$
1	1	1	1	8
2	0	0	0	0
3	1	0	0	4
4	0	1	0	4

Discrimination rule example: $\text{Spam}(x) = 4x_1 + 4x_2 > 5$

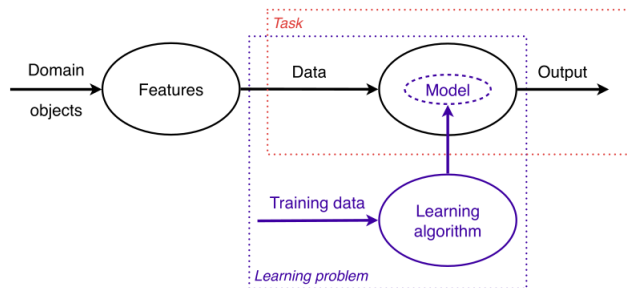
Definizione 1.1.1: Apprendimento automatico

L'apprendimento automatico è lo studio sistematico di algoritmi e sistemi che migliorano le loro conoscenze e performance con l'esperienza.

L'apprendimento automatico è interessato a usare le giuste features per costruire il giusto modello per ottenere buone performance sul giusto task.

Domanda 1.1

L'apprendimento automatico come può aiutarci a risolvere un task?



Dal dominio dell'applicazione arrivano degli oggetti descritti tramite features che vengono utilizzate per creare dei *training data* e un *dataset*. Questi vengono usati per costruire un modello per calcolare un output.

Note:-

Per risolvere un task bisogna sfruttare un modello. Per risolvere un problema di apprendimento bisogna trovare un algoritmo di apprendimento.

1.1.1 Tasks

Definizione 1.1.2: Tasks predittivi

Un task predittivo è focalizzato sul predire una variabile sulla base degli esempi. Si parte da problemi vecchi per trovare la soluzione a *nuovi* problemi.

Corollario 1.1.1 Overfitting

L'Overfitting è un adattamento eccessivo al dataset di allenamento per cui, messi di fronte a nuovi problemi, non si riesce a trovare una soluzione soddisfacente.

I tasks predittivi possono essere:

- *binari e multi-classe*: di categorizzazione;
- *Regressivi*: con un target numerico;
- *Clustering*: un target sconosciuto.

Note:-

IL Clustering fa anche parte dei tasks descrittivi.

Definizione 1.1.3: Tasks descrittivi

Un task descrittivo si concentra sul fornire regolarità nel dataset.

	Films			
Users	1	0	1	0
	0	2	2	2
	0	0	0	1
	1	2	3	2
	1	0	1	1
	0	2	2	3
	The Shawshank Redemption	The Usual Suspects	The Godfather	The Big Lebowski

Questa matrice rappresenta i voti dati da utenti a dei film. Si vogliono estrapolare le caratteristiche di questi film che hanno generato questi voti. Guardando questa matrice individualmente è difficile, per cui si compone con altre matrici.

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 2 & 2 \\ 0 & 0 & 0 & 1 \\ 1 & 2 & 3 & 2 \\ 1 & 0 & 1 & 1 \\ 0 & 2 & 2 & 3 \end{pmatrix} = \begin{matrix} \text{Users} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \end{matrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{matrix} \text{Films} \\ \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

drama
crime
comedy
The Shawshank Redemption
The Usual Suspects
The Godfather
The Big Lebowski

1.1.2 Modelli

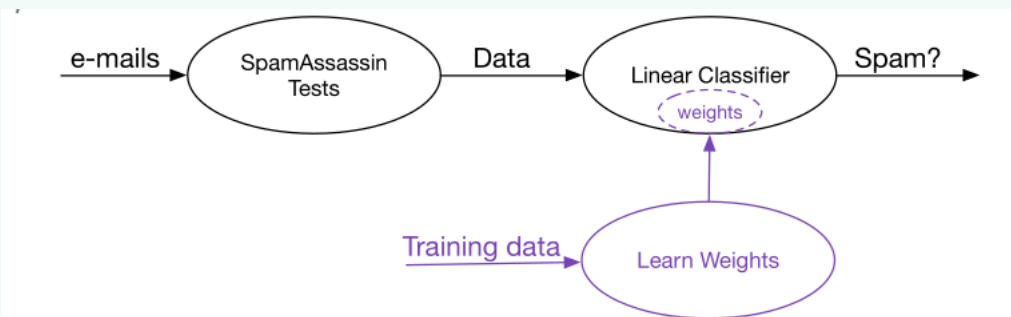
Ci sono 3 possibili tipi di modelli:

- *Geometrici*: modelli che usano l'intuizione dalla geometria per risolvere il problema;
- *Probabilistici*: usano il calcolo delle probabilità;
- *Logici*.

Definizione 1.1.4: Modelli geometrici

Nei modelli geometrici gli esempi sono punti di uno spazio vettoriale e la loro classificazione corrisponde a trovare un iperpiano che separi i punti positivi da quelli negativi.

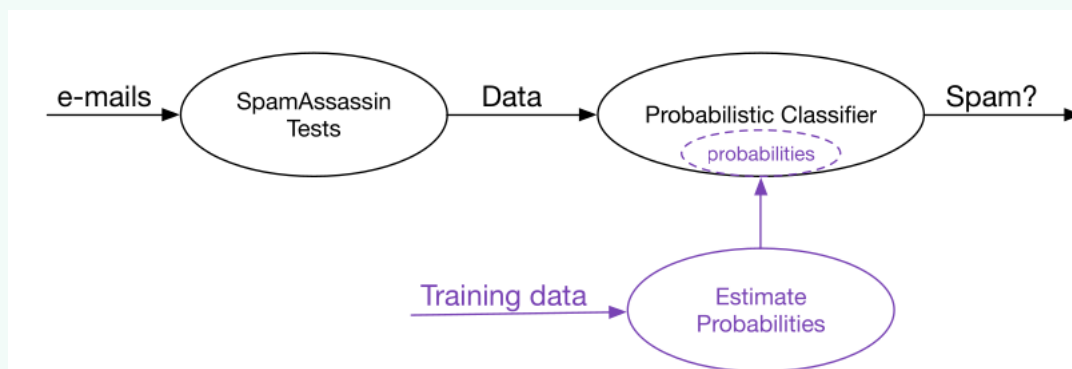
Esempio 1.1.2 (Modello geometrico)



Definizione 1.1.5: Modelli probabilistici

Nei modelli probabilistici si fanno delle stime con dei classificatori probabilistici. Dopo di che si usano delle regole di decisione.

Esempio 1.1.3 (Modello probabilistico)

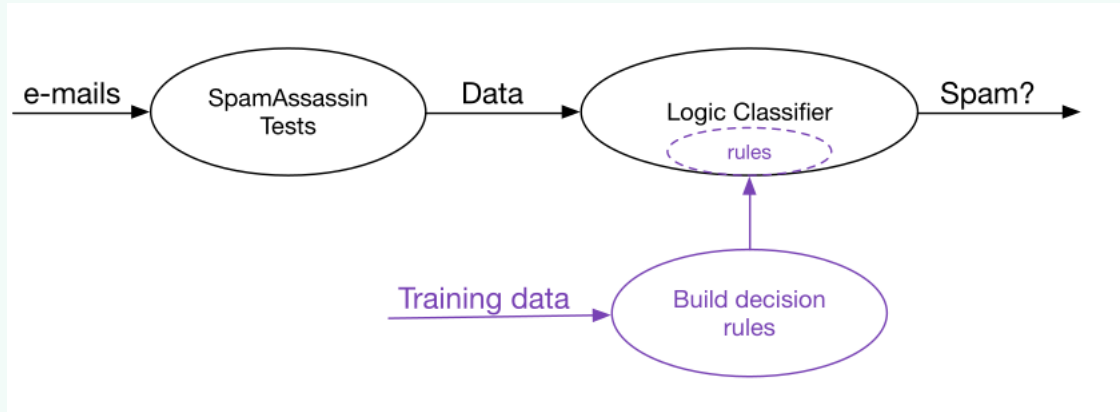


Note:-

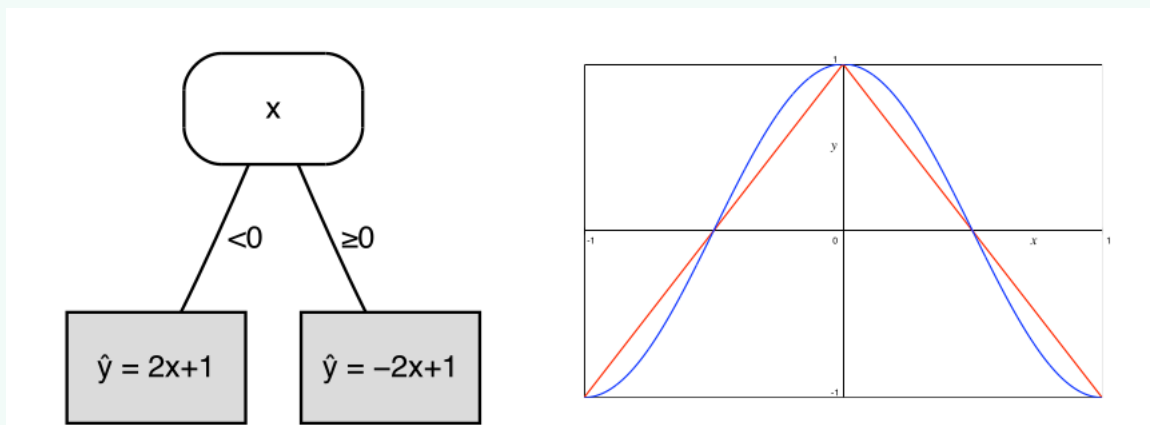
Uno degli algoritmi più semplici che si utilizza con i modelli probabilistici è l'assunzione di Naive Bayes. Si assume che x_1 e x_2 siano indipendenti tra loro per cui si possono calcolare solo i valori di x_1 e di x_2 individualmente.

Definizione 1.1.6: Modelli logici

Nei modelli logici si utilizza la logica. Si hanno una serie di regole.

Esempio 1.1.4 (Modello logico)**1.1.3 Features****Definizione 1.1.7: Features**

Il modo in cui si descrivono i propri dati. Possono facilitare il lavoro di apprendimento se correttamente usate.

Esempio 1.1.5 (Coseno)

Due rappresentazioni della funzione coseno: a sinistra si utilizza una variabile di regressione, a destra un'approssimazione lineare.

2

Test2

