ANNO ACCADEMICO 2024/2025

Apprendimento Automatico

Teoria

Luna's Notes



DIPARTIMENTO DI INFORMATICA

CAPITOLO 1	Introduzione	Pagina 5
1.1	Le basi del machine learning	5
Capitolo 2	Test2	Pagina 8

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 contente 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 definizione

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:

- \Rightarrow *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)

```
Test scores: { w<sub>i</sub> }
                                                                                                                                                 RBL: MXRate recommends allowing [123.45.6.789 listed in sub.mxrate.net] BODY: HTML has a low ratio of text to image area
-0.1 RCVD_IN_MXRATE_WL
 0.6 HTML_IMAGE_RATIO_02
1.2 TVD_FW_GRAPHIC_NAME_MID
0.0 HTML_MESSAGE
0.6 HTML_FONx_FACE_BAD

BODY: TVD_FW_GRAPHIC_NAME_MID
BODY: HTML included in message
BODY: HTML font face is not a verified by the second s
                                                                                                                                                  BODY: HTML font face is not a word
1.4 SARE_GIF_ATTACH
0.1 BOUNCE_MESSAGE
                                                                                                                                                  FULL: Email has a inline gif
                                                                                                                                                  MTA bounce message
 0.1 ANY_BOUNCE_MESSAGE
                                                                                                                                                  Message is some kind of bounce message
 1.4 AWL
                                                                                                                                                  AWL: From: address is in the auto white-list
                                                             Test ID
                                                                                                                                                                                                                                             Test Description
           \sum w_i > 5 \Rightarrow the message is likely to be spam
```

Spam
Assassin 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: 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.

Test2