

Appunti machine learning

Marco

June 28, 2021

Contents

1	Termini base	1
1.1	Modello	1
1.2	Dataset in input	1
1.3	Supervised learning	2
1.4	Unsupervised learning	2
1.5	Etichetta	2
1.6	Cluster	2
2	Prerequisiti	2

1 Termini base

Per poter effettuare delle previsioni su un dato, è necessario definire alcuni componenti:

1.1 Modello

Relazione tra le caratteristiche e l'output. Rappresentazione interna dei dati che l'algoritmo ha appreso finora. Scatola nera che dati dei valori in input, fornisce una previsione in base a ciò che ha appreso finora.

1.2 Dataset in input

Per poter effettuare il training del modello, è necessario creare un insieme di dati da fornire in input. Questi dati sono costituiti da un insieme di caratteristiche (ovvero dei parametri che definiscono il dato) e, nel caso di supervised learning, da un'etichetta, che rappresenta il dato di output. Nel

caso di unsupervised learning, l'etichetta non viene fornita dal data scientist ma deve essere dedotta dall'algoritmo di learning.

1.3 Supervised learning

Il data scientist fornisce l'etichetta insieme ai dati.

1.4 Unsupervised learning

Il data scientist non fornisce l'etichetta insieme ai dati. L'algoritmo di machine learning deve ricavare anche l'etichetta.

1.5 Etichetta

Valore in output a cui corrispondono le varie caratteristiche che caratterizzano il dato.

1.6 Cluster

Per poter effettuare previsioni su dei dati non etichettati, è necessario definire dei cluster, ovvero un gruppo di dati. Il numero di cluster deve essere uguale al numero di possibili etichette. L'algoritmo di learning definisce i cluster in base a dei possibili criteri matematici: ad esempio la distanza relativa tra ogni punto. Per esempio, per fornire una previsione su un nuovo dato, il modello prende la distanza del nuovo dato da ogni cluster e sceglie il cluster più vicino.

2 Prerequisiti