

Tarallucci, Vino e Machine Learning
Corso “Il paper della buonanotte”

L'impostazione Statistica del Machine Learning

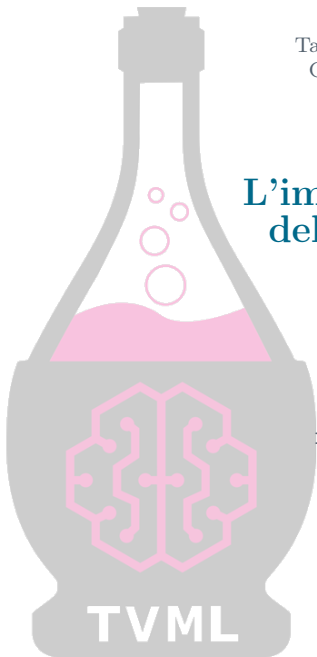
Parte I

Fabio Mardero

fabio.mardero@gmail.com

github.com/fmardero

4 aprile 2019





Il Machine Learning

Apprendimento

Compito

Analisi Matematica Unidimensionale

Spazio topologico

Spazio metrico

La derivata

L'integrale

Probabilità e Statistica



Il Machine Learning



Il machine learning è un gruppo di modelli matematici in grado di “apprendere” dai dati allo scopo di eseguire, nel modo migliore possibile, un dato compito.

Caratterizzazione di un modello di ML

Un modello di machine learning è quindi caratterizzato da

- ▶ un apprendimento
- ▶ un compito



In termini matematici, un modello apprende quando modifica la sua struttura, o i suoi parametri, per ridurre gli errori delle sue previsioni.

Può essere paragonato ad un agente collocato in un dato ambiente (*environment*). È esattamente ciò che accade per un algoritmo di ML messo in produzione.

L'algoritmo può interagire con l'ambiente o “subirlo”.



L'ambiente e/o l'interazione con esso produce un fenomeno i cui effetti misurabili sono raccolti come dati.

I dati possono essere

- ▶ strutturati, organizzati in database detti *dataset*,
- ▶ non strutturati, conservati senza alcuno schema,
- ▶ semi-strutturati.



Metodi di apprendimento

L'apprendimento di un modello di machine learning può avvenire in tre diversi modi:

- ▶ per rinforzo (*reinforcement learning*),
- ▶ in maniera supervisionata (*supervised learning*),
- ▶ in maniera non supervisionata (*unsupervised learning*).



Italiano

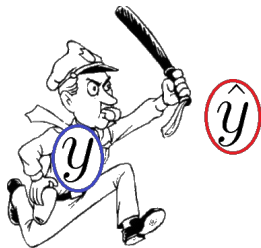
L'agente interagisce con l'environment e ogni sua azione modifica l'ambiente stesso.

Matematiche

Il modello interagisce con il sistema e ogni sua previsione modifica lo stato dello stesso.



Nel tempo, non necessariamente ad ogni interazione con l'ambiente, l'agente riceve un *feedback* sul suo comportamento. Egli modifica quindi le sue future azioni, sulla base delle precedenti, tentando di massimizzare quelle che hanno portato a risultati positivi e minimizzando quelle risultate negative. L'apprendimento dipende quindi da un sistema di *rewards* e *punishments*.



Il modello subisce l'ambiente.
Nel caso dell'apprendimento supervisionato il modello mira a predire il comportamento di una o più variabili osservate rispetto alle altre.

Indicata con \hat{y} la previsione e con y il valore osservato, il modello apprende a minimizzare l'errore tra \hat{y} e y . L'apprendimento è, informalmente, “supervisionato” dai valori di y .



Il modello subisce l'ambiente ma non è allenato per fornire una previsione.

L'apprendimento non supervisionato prevede che l'algoritmo ricerchi strutture informative (*pattern*) tra i dati.



Il compito definisce su cosa il modello è allenato e con quali intenzioni. L'oggetto di analisi sono dati strutturati o semi-strutturati.

Si riconoscono due casi:

1. si individuano delle variabili più importanti, dette **variabili target**/risposta, rispetto alle altre, chiamate **variabili esplicative**/covariate/features,
2. tutte le variabili sono intese come significative (o potenzialmente tali).

Dato un insieme di dati, spetta all'osservatore decidere come intende interpretarli e se assegnare particolare importanza a qualcuna delle variabili disponibili.



Caso 1

Compiti di regressione o classificazione.

Mirando a fornire una previsione accurata delle variabili target, il modello spiega il fenomeno che genera y .

Caso 2

Compiti legati all'estrazione di informazione dai dati e ad una loro rappresentazione, ad esempio il clustering.

Ad esempio si individuano somiglianze tra informazioni presenti nel dataset.

Il modello di machine learning, a discapito del compito, fornisce un'interpretazione del fenomeno che genera i dati. Cambia la finalità esplicativa.



Analisi Matematica Unidimensionale



Fonte:

Gianluca Occhetta - *“Note di TOPOLOGIA GENERALE e primi elementi di topologia algebrica”*



Dato un generico insieme A , si definisce **topologia** τ la famiglia dei sottoinsiemi di A tale che

- ▶ $\emptyset, A \in \tau$
- ▶ La famiglia τ è chiusa rispetto all'unione

$$\text{data } \{U_i\}_{i \in I} \text{ con } U_i \in \tau \implies \bigcup_i U_i \in \tau$$

- ▶ La famiglia τ è chiusa rispetto alle intersezioni finite

$$\text{se } U_i, U_j \in \tau \implies U_i \cap U_j \in \tau$$

Lo **spazio topologico** è la coppia (A, τ) dove A è un insieme e τ una topologia. Gli insiemi $U \in \tau$ si dicono **insiemi aperti**.



Si definisce insieme delle parti di A , indicato con $\mathcal{P}(A)$, l'insieme di tutti i sottoinsiemi di A . L'insieme delle parti è quindi una topologia su A , detta *topologia discreta*.

Esempio

$$A = \{a, b\}$$

allora

$$\tau = \mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, A\}$$

La topologia non è unica in quanto si può definire ad esempio la *topologia banale*

$$\tau = \{\emptyset, A\}$$



Intorno

Sia $x \in A$; un intorno (aperto) di x è un sottoinsieme $I(x) \subset A$ tale che contiene un insieme aperto U che include x .

$$x \in U \subset I(x) \subset A$$

Interno di un insieme

Sia $U \in A$; un punto $x \in U$ si dice interno a U se esiste un intorno $I(x)$ tale che $I(x) \subset U$. L'insieme di tutti i punti interni di U è detto interno di U e si denota con \mathring{U} . Si osservi che U è aperto se e solo se $U = \mathring{U}$.



Chiusura di un insieme

Sia $U \in A$; un punto $x \in A$ è di aderenza per U se per ogni intorno $I(x)$ di x si ha che $I(x) \cap U \neq \emptyset$. L'insieme di tutti i punti di aderenza di U in A è detto chiusura di U e si denota con \bar{U} .

Frontiera di un insieme

La frontiera di U , indicata con ∂U , è l'insieme dato da $\bar{U} \setminus \mathring{U}$.



Inserisci immagine esplicativa.



Punto di accumulazione

Sia $U \in A$; un punto $x_0 \in A$ (non deve necessariamente appartenere a U) si dice di accumulazione per U se per ogni intorno $I(x_0)$ esiste almeno un elemento x tale che $x \neq x_0$ e $x \in U$.

$$\forall I(x_0) \quad \exists x \in U : x \in I(x_0) \setminus \{x_0\}$$

Intuitivamente significa che a qualsiasi livello di ingrandimento attorno a x_0 si continuano a vedere punti di U (diversi da x_0).

“Sai dirmi un numero x positivo vicino a $x_0 = 0$ tale per cui non tra x e x_0 non ci sono altri numeri?”



Siano (A, τ) e (B, σ) due spazio topologici, f una funzione

$$f : A \longrightarrow B$$

e $x_0 \in A$ punto di accumulazione. Si definisce **limite** $L \in B$ **della funzione f per $x \in A$ che tende al punto x_0** , indicato con

$$L = \lim_{x \rightarrow x_0} f(x)$$

il valore tale per cui

$$\forall V \in \{I(L)\} \quad \exists U \in \{I(x_0)\} : \quad f(U \setminus \{x_0\}) \subset V$$

indicando con $\{I(L)\}$ e $\{I(x_0)\}$ la famiglie di intorni definiti rispettivamente a L e x_0 .



Funzione continua

Siano (A, τ) e (B, σ) due spazi topologici. Una funzione f

$$f : A \longrightarrow B$$

si dice **continua** se la controimmagine di ogni insieme aperto di B è un aperto di A , cioè se

$$\forall V \in \sigma \implies f^{-1}(V) \in \tau$$

La continuità di una funzione dipende non solo dagli insiemi A e B , ma anche dalle topologie su di essi considerate.



Omeomorfismo

Due spazi topologici (A, τ) e (B, σ) si dicono **omeomorfi** se esistono due funzioni continue f e g

$$f : A \longrightarrow B$$

$$g : B \longrightarrow A$$

tali che $g \circ f = \mathbb{1}_A$ e $f \circ g = \mathbb{1}_B$. Le due funzioni si dicono omeomorfismi e sono quindi continue, biunivoche e con inversa continua.

L'idea di omeomorfismo permette di formalizzare l'idea che per passare da A e B , e viceversa, basta deformare lo spazio senza “strappi”. Per un esempio si veda: from cup to toro.



Sia (A, τ) uno spazio topologico.

Si definisce $\mathcal{B} \subset \tau$ **base della topologia** τ il sottoinsieme di τ tale per cui ogni aperto non vuoto $U \in \tau$ è unione di elementi di \mathcal{B} .

Fissato $U \in \tau$, $U \neq \emptyset$

$$U = \bigcup_i B_i \quad B_i \in \mathcal{B} \subset \tau$$

Informalmente

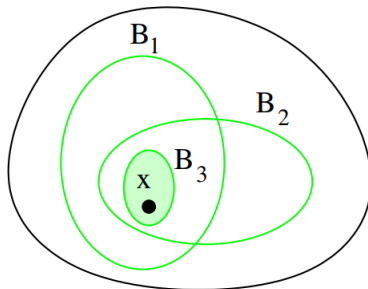
Con alcuni particolari elementi della topologia, che compongono la base, sono in grado di “costruire” qualsiasi insieme contenuto in τ .



Teorema di caratterizzazione delle basi

Se \mathcal{B} è una base di una topologia τ su A allora

- ▶ $\forall x \in A \quad \exists B \in \mathcal{B}$ tale che $x \in B$
- ▶ $\forall B_1, B_2 \in \mathcal{B}$ tale che $B_1 \cap B_2 \neq \emptyset$ e $\forall x \in B_1 \cap B_2$ allora $\exists B_3 \in \mathcal{B}$ tale che $x \in B_3 \subset B_1 \cap B_2$





Vale anche viceversa.

Dato un insieme A e una famiglia di sottoinsiemi \mathcal{B} che soddisfa le due proprietà, allora esiste un'unica topologia τ su A che ha \mathcal{B} come base.

Informalmente

Trovato un sottoinsieme \mathcal{B} di A che rispetta le proprietà sopracitate allora si identifica automaticamente una topologia τ con base \mathcal{B} .



aggiungere più e meno infinito *Assioma di completezza* Dati a, b qualsiasi per cui vale $a \leq b$ allora $\exists c$ tale che $a \leq c \leq b$.



Si definisce su \mathbb{R} un intervallo aperto come

$$]a, b[= \{x \in \mathbb{R} : a < x < b\}$$

con $a \in \mathbb{R} \cup \{-\infty\}$ e $b \in \mathbb{R} \cup \{+\infty\}$.

Si trova allora che, dato $A \subset \mathbb{R}$,

$$\forall x \in A \quad \exists]a, b[\quad \text{tale che} \quad x \in]a, b[\subset A$$



Dato un generico insieme A , si definisce **distanza** d su A una funzione

$$d : A \times A \longrightarrow \mathbb{R}$$

tale che

► *Positività*

$$d(x, y) \geq 0 \quad \forall x, y \in A \quad \text{e} \quad d(x, y) = 0 \iff x = y$$

► *Simmetria*

$$d(x, y) = d(y, x) \quad \forall x, y \in A$$

► *Disuguaglianza triangolare*

$$d(x, y) \leq d(x, z) + d(z, y) \quad \forall x, y, z \in A$$

Lo **spazio metrico** è la coppia (A, d) dove A è un insieme e d una distanza.



La distanza induce una topologia nello spazio metrico, e per dimostrarlo basta trovare una famiglia di insiemi che rispetti le due proprietà. Per ogni $x_0 \in A$ e $r > 0$, si definisce la **palla**

$$B_{x_0, r} = \{x \in A : d(x, x_0) < r\}$$

così che $\mathcal{B} = \{B_{x_0, r}; x_0 \in A, r > 0\}$. Sfruttando le proprietà della distanza risulta immediato dimostrare la validità delle due proprietà.

Si osservi che la palla $B_{x_0, r}$ è un intorno di x_0 .



La norma come caso particolare di distanza. Elencare relative proprietà



Esempi di distanze

Distanza euclidea

$$]a, b[= \{x \in \mathbb{R} : a < x < b\}$$



Dato un generico insieme A , sia f una funzione

$$f : A \longrightarrow \mathbb{R}$$

e $x_0 \in A$ punto di accumulazione. Si definisce **limite L della funzione f per $x \in A$ che tende al punto x_0** , indicato con

$$L = \lim_{x \rightarrow x_0} f(x)$$

il valore tale per cui

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad 0 < d(x, x_0) < \delta \quad \implies 0 < d(f(x), L) < \varepsilon$$

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad 0 < |x - x_0| < \delta \quad \implies 0 < |f(x) - L| < \varepsilon$$



Nella definizione di limite nello spazio reale ritorna ovviamente il tema della distorsione senza strappi. Esempi di continuità o meno.



Sia

$$f : A \subset \mathbb{R} \longrightarrow \mathbb{R}$$

allora la funzione è continua nel punto $x_0 \in A$ se e solo se

$$\lim_{x \rightarrow x_0} f(x) = f(x_0)$$

cioè

$$\lim_{x \rightarrow x_0^-} f(x) = \lim_{x \rightarrow x_0^+} f(x) = f(x_0)$$

Se la relazione vale per ogni punto x_0 del dominio, allora la funzione si dice essere continua su A .



Si presti attenzione che nella definizione di f si è fatto uso della retta reale e non di quella ampliata.

Il calcolo dei limiti di funzioni continue risulta molto semplice e segue direttamente da...

Esempio della theta di Heaviside e della sigmoide.



Sia

$$f : A \subset \mathbb{R} \longrightarrow \mathbb{R}$$

allora, sotto opportune ipotesi, la derivata di f rispetto ad un punto $x_0 \in A$ si scrive come limite del *rapporto incrementale*

$$f'(x_0) = \frac{df(x_0)}{dx} = \left. \frac{df(x)}{dx} \right|_{x=x_0} = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$



Per alcune particolari funzioni il calcolo della derivata risulta estremamente semplice.

Proprietà della derivata come approssimante lineare, tangente alla curva e direzione di massima pendenza

Ripasso di Analisi Matematica Unidim.

Teorema di Taylor



Ripasso di Analisi Matematica Unidim.

L'integrale





Probabilità e Statistica





Variabili aleatorie discrete e continue.



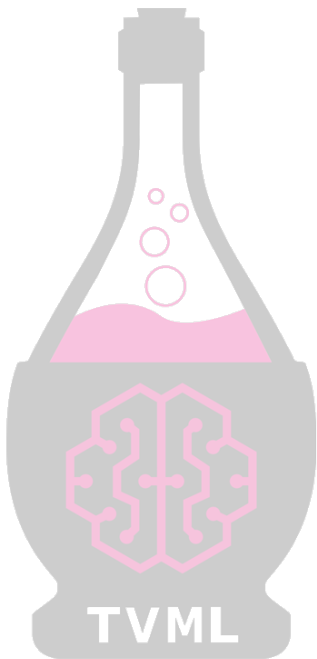
Partizione dell'evento certo Ω .



Per una variabile aleatoria continua X , la probabilità che si verifichi un dato evento $\{X = x_0\}$ è nulla.

$$P(X = x_0) = \int_{x_0}^{x_0} f(x) dx = 0$$

Generalmente è più semplice costruire modelli matematici nel continuo, che sfruttano le proprietà degli integrali, per poi facilmente ricondurre i risultati a variabili aleatorie discrete.



Grazie dell'attenzione!