

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

L'impostazione Statistica del Machine Learning

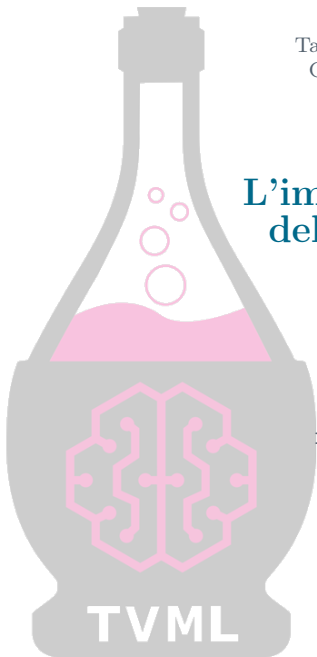
Parte I

Fabio Mardero

fabio.mardero@gmail.com

github.com/fmardero

4 aprile 2019





Il Machine Learning

Apprendimento

Compito

Topologia

Spazio Topologico

La Retta Reale

Spazio Metrico

Spazio Normato

Analisi Matematica Unidimensionale

Il limite

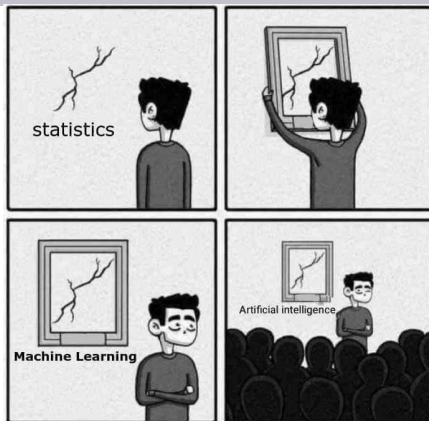
La derivata

Teoria della Misura

Spazio Misurabile

L'integrale

Probabilità





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

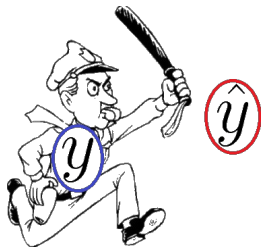
L'agente interagisce con l'environment e ogni sua azione modifica l'ambiente stesso.

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.



Topologia



Fonte:

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



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

- ▶ $\emptyset, \Omega \in \tau$
- ▶ 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 (Ω, 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 Ω , indicato con $\mathcal{P}(\Omega)$, l'insieme di tutti i sottoinsiemi di A . L'insieme delle parti è quindi una topologia su Ω , detta *topologia discreta*.

Esempio

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

allora

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

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

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



Intorno

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

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

Interno di un insieme

Sia $U \in \Omega$; 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 $U = \mathring{U}$.

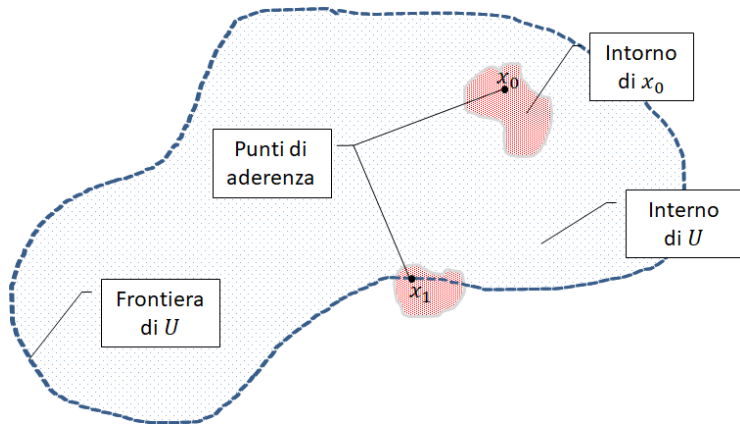


Chiusura di un insieme

Sia $U \in \Omega$; un punto $x \in \Omega$ 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 Ω 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 $\bar{U} \setminus \mathring{U}$.





Punto di accumulazione

Sia $U \in \Omega$; un punto $x_0 \in \Omega$ (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 (Ω, τ) e (B, σ) due spazio topologici, f una funzione

$$f : \Omega \longrightarrow B$$

e $x_0 \in \Omega$ punto di accumulazione. Si definisce **limite** $L \in B$ **della funzione f per $x \in \Omega$ 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 .



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

$$f : \Omega \longrightarrow B$$

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

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

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



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

$$f : \Omega \longrightarrow B$$

$$g : B \longrightarrow \Omega$$

tali che $g \circ f = \mathbb{1}_\Omega$ 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 Ω e B , e viceversa, basta deformare lo spazio senza “strappi”. Per un esempio si veda: from cup to toro.



Sia (Ω, 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 $U \in \tau, U \neq \emptyset$

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

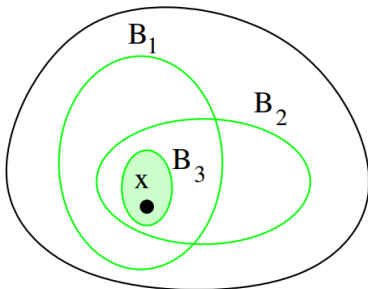
Informalmente

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 Ω allora

- ▶ $\forall x \in \Omega \quad \exists B \in \mathcal{B}$ 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 Ω e una famiglia di sottoinsiemi \mathcal{B} che soddisfa le due proprietà, allora esiste un'unica topologia τ su Ω che ha \mathcal{B} come base.

Informalmente

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



La retta reale è definita come l'insieme dei numeri che soddisfano le proprietà di campo, comuni a \mathbb{Q} , e l'assioma di completezza.

Assioma di completezza

Dati $a, b \in \mathbb{R}$ qualsiasi per cui vale $a \leq b$ allora $\exists c \in \mathbb{R}$ tale che $a \leq c \leq b$.

Si parla di **retta reale ampliata** l'insieme di punti definito come $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, +\infty\}$ tale che

$$-\infty < x < +\infty \quad \forall x \in \mathbb{R}$$



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 $\Omega \subset \mathbb{R}$,

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

Si può dimostrare che **intervalli aperti sono insiemi aperti e definiscono una topologia su \mathbb{R} .**



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

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

tale che

► *Positività*

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

► *Simmetria*

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

► *Disuguaglianza triangolare*

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

Lo **spazio metrico** è la coppia (Ω, 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 $x_0 \in \Omega$ e $r > 0$, si definisce la **palla**

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

così che $\mathcal{B} = \{B_{x_0, r}; x_0 \in \Omega, r > 0\}$. 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 .

Esempi di distanze

- *Distanza discreta*

$$d(x, y) = \begin{cases} 0, & x = y \\ 1, & x \neq y \end{cases}$$

- *Distanza indotta dalla norma*

$$d(x, y) = \|x - y\|$$

- *Distanza dell'estremo superiore*

$$d(f, g) = \sup_{x \in A} \|f(x) - g(x)\|$$



Si definisce **norma** su uno spazio vettoriale Ω la funzione

$$\|\cdot\| : \Omega \longrightarrow [0, +\infty[$$

tale che rispetta le seguenti proprietà

► *Positività*

$$\|x\| \geq 0 \quad \forall x \in \Omega \quad \text{e} \quad \|x\| = 0 \iff x = 0$$

► *Omogeneità*

$$\|\lambda x\| = |\lambda| \|x\| \quad \forall x \in \Omega, \lambda \in \mathbb{R}$$

► *Disuguaglianza triangolare*

$$\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in \Omega$$

Lo **spazio normato** è la coppia $(\Omega, \|\cdot\|)$ dove Ω è un insieme e $\|\cdot\|$ una norma.



Su uno spazio \mathbb{R}^n la norma può essere definita come

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$$

con $p \in [1, +\infty[$.

Con $p = 1$ si definisce la distanza di Manhattan, con $p = 2$ la consueta distanza euclidea.

Per le proprietà che la caratterizzano, la norma induce su uno spazio metrico una distanza invariante per traslazioni.

La norma è usualmente definita come il risultato, sotto radice, del prodotto scalare

$$\|x\| = \sqrt{x \cdot x}$$



Analisi Matematica Unidimensionale



La definizione di limite si può ora concretizzare maggiormente.
Dato un generico insieme $A \subset \mathbb{R}$, sia f una funzione

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

e x_0 punto di accumulazione per A . 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

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

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



Il limite di f in x_0 esiste se e solo se esistono il limite destro e sinistro e coincidono

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

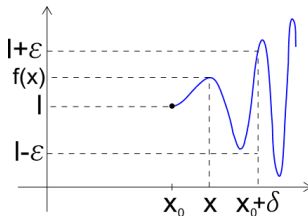
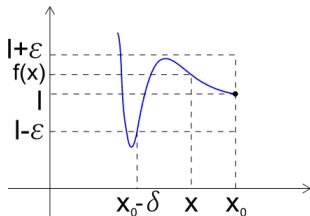


Figura: youmath.it

Nel calcolo dei limiti si può utilizzare anche la retta reale ampliata. Esistono “limiti notevoli”.



Sia

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

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

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

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



Esempi

- *Retta (generica)*

$$f(x) = mx + q$$

$$\lim_{x \rightarrow -\infty} f(x) = \text{sign}(m) (-\infty) \qquad \lim_{x \rightarrow +\infty} f(x) = \text{sign}(m) (+\infty)$$



Esempi

► *Tangente Iperbolica*

$$f(x) = \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\lim_{x \rightarrow -\infty} f(x) = -1 \qquad \lim_{x \rightarrow +\infty} f(x) = +1$$

► *Theta di Heaviside*

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

$$\lim_{x \rightarrow 0^-} f(x) = 0 \qquad \lim_{x \rightarrow 0^+} f(x) = 1$$



Sia A aperto e

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

allora, se il limite esiste ed è finito, la derivata di f rispetto ad un punto $x_0 \in A$ si scrive come limite del *rapporto incrementale*

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

Funzioni derivabili sono anche continue, non vale necessariamente viceversa.



Alcune regole di derivazione

- **Constant Rule:** $f(x) = c$ then $f'(x) = 0$
- **Constant Multiple Rule:** $g(x) = c \cdot f(x)$ then $g'(x) = c \cdot f'(x)$
- **Power Rule:** $f(x) = x^n$ then $f'(x) = nx^{n-1}$
- **Sum and Difference Rule:** $h(x) = f(x) \pm g(x)$ then $h'(x) = f'(x) \pm g'(x)$
- **Product Rule:** $h(x) = f(x)g(x)$ then $h'(x) = f'(x)g(x) + f(x)g'(x)$
- **Quotient Rule:** $h(x) = \frac{f(x)}{g(x)}$ then $h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$
- **Chain Rule:** $h(x) = f(g(x))$ then $h'(x) = f'(g(x))g'(x)$



Esempi

► Funzione Sigmoide

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x))$$

► ReLU

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$

$$\nexists f'(0)$$

infatti

$$\lim_{h \rightarrow 0^-} \frac{f(h) - f(0)}{h} = 0 \qquad \lim_{h \rightarrow 0^+} \frac{f(h) - f(0)}{h} = 1$$



La derivata di f calcolata in x_0 è il coefficiente angolare della retta tangente alla curva $f(x)$ passante per il punto $(x_0, f(x_0))$. La retta tangente in x_0 ha equazione

$$y = f(x_0) + f'(x_0)(x - x_0)$$

ed è chiamata **approssimante lineare** poiché

$$m = f'(x_0)$$

$$q = f(x_0) - f'(x_0) x_0$$

Nella Parte II si mostrerà che la derivata, o più in generale il gradiente, è la direzione di massima pendenza della funzione nel punto x_0 .



Teorema di Taylor

Sia $]a, b[\subset \mathbb{R}$ un intervallo aperto, $x_0 \in]a, b[$ e $f: (a, b) \rightarrow \mathbb{R}$ derivabile $(n-1) \geq 0$ volte nell'intervallo. Si suppone che la derivata n -esima $f^{(n)}$ sia continua nel punto x_0 . Allora la funzione f in x può essere scritta come

$$f(x) = \underbrace{\sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k}_{\text{polinomio di Taylor}} + \underbrace{o((x - x_0)^n)}_{\text{infinitesimo di ordine superiore a } (x - x_0)^n}$$

dove

$$\lim_{x \rightarrow x_0} \frac{o((x - x_0)^n)}{(x - x_0)^n} = 0$$



Teoria della Misura



Fonte:

D. Bertacchi, M.U. Dini - *“Compendio di teoria della misura (con un occhio alla probabilità)”*



Dato un generico insieme Ω , si definisce σ - **algebra** \mathcal{A} la famiglia dei sottoinsiemi di Ω tale che

- ▶ $\emptyset, \Omega \in \mathcal{A}$
- ▶ *La famiglia \mathcal{A} è chiusa rispetto alla formazione di complementari*

$$\text{se } U \in \mathcal{A} \implies U^c \in \mathcal{A}$$

- ▶ *La famiglia \mathcal{A} è chiusa rispetto alla formazione di unioni numerabili*

$$\{U_n\}_{n \in \mathbb{N}} \subset \mathcal{A} \implies \bigcup_n U_n \in \mathcal{A}.$$

Lo **spazio misurabile** è la coppia (Ω, \mathcal{A}) dove Ω è un insieme e \mathcal{A} una σ - algebra. Gli insiemi $U \in \mathcal{A}$ si dicono **insiemi \mathcal{A} -misurabili**.



Data una famiglia \mathcal{F} di insiemi su Ω , allora

$$\sigma(\mathcal{F}) = \bigcap_{\substack{\mathcal{F}_\alpha \supseteq \mathcal{F} \\ \mathcal{F}_\alpha \text{ } \sigma\text{-algebra}}} \mathcal{F}_\alpha$$

è una σ -algebra generata da \mathcal{F} (la più piccola che contenga \mathcal{F}).



Dato uno spazio topologico (Ω, τ) , essendo τ una famiglia di sottoinsiemi di Ω , allora

$$\sigma(\tau) = \bigcap_{\substack{\tau_\alpha \supseteq \tau \\ \tau_\alpha \text{ } \sigma\text{-algebra}}} \tau_\alpha$$

è una σ -algebra, detta **σ -algebra di Borel**.

Gli elementi di $\sigma(\tau)$ si dicono **boreliani**.

Se $\Omega = \mathbb{R}^n$ allora la relativa σ -algebra di Borel sarà indicata con \mathcal{B}^n (per $\Omega = \mathbb{R}$ si userà \mathcal{B}^1).



Data una σ -algebra \mathcal{A} , una funzione μ

$$\mu: \mathcal{A} \rightarrow [0, +\infty]$$

si dice **misura di \mathcal{A}** se soddisfa la *proprietà di σ -additività*, cioè data $\{U_n\}_{n \geq 1}$ successione di insiemi \mathcal{A} -misurabili a due a due disgiunti

$$\mu \left(\sum_{n=1}^{+\infty} U_n \right) = \sum_{n=1}^{+\infty} \mu(U_n)$$



- ▶ Se $\mu(\Omega) < +\infty$ allora si parla di *misura finita*.
- ▶ Se $\mu(\Omega) = 1$ allora si parla di *misura di probabilità* $\mu = P$. In questo caso gli insiemi misurabili sono detti **eventi**.

La terna $(\Omega, \mathcal{A}, \mu)$ è detta **spazio di misura** (o **spazio di probabilità** se $\mu = P$ è misura di probabilità).



Alcune proprietà della misura

- ▶ $\mu(\emptyset) = 0$
- ▶ μ è *finitamente additiva*.

$$\mu\left(\sum_{n=1}^n U_n\right) = \sum_{n=1}^n \mu(U_n)$$

- ▶ μ è *monotona*.

$$A \subseteq B \implies \mu(A) \leq \mu(B)$$



Misura di conteggio

La misura di conteggio ν_C è definita su $\mathcal{P}(\Omega)$ come

$$\nu_C(A) = \begin{cases} \#A, & A \text{ finito} \\ +\infty, & A \text{ infinito} \end{cases}$$

con $\#A$ numero di elementi dell'insieme.



Dato (Ω, \mathcal{A}) e (Θ, \mathcal{B}) due spazi misurabili, una funzione $f: \Omega \rightarrow \Theta$ si dice misurabile o $(\mathcal{A}, \mathcal{B})$ -misurabile se

$$f^{-1}(V) = \{f \in V\} = \{\omega \in \Omega : f(\omega) \in V\} \in \mathcal{A} \quad \forall V \in \mathcal{B}$$

Se $\Theta = \mathbb{R}$ e $\mathcal{B} = \mathcal{B}^1$ allora f si dice *funzione di Borel*.



Lemma di misurabilità di funzioni continue

Siano (Ω, τ) e (Θ, σ) due spazi topologici, e siano (Ω, \mathcal{A}) e (Θ, \mathcal{B}) i relativi spazi boreliani. Se una funzione $f : \Omega \mapsto \Theta$ è continua rispetto a τ e σ allora essa è anche $(\mathcal{A}, \mathcal{B})$ -misurabile.



Dato uno spazio di misura $(\Omega, \mathcal{A}, \mu)$, sia s una funzione misurabile semplice positiva

$$s = \sum_{i=1}^n \alpha_i I(A_i)$$

con $I(A_i)$ la funzione indicatrice.

L'integrale di s su $E \in \mathcal{A}$ si definisce come

$$\int_E s(\omega) d\mu(\omega) = \sum_{i=1}^n \alpha_i \mu(A_i \cap E)$$



Data f funzione misurabile positiva

$$f: \Omega \rightarrow [0, +\infty]$$

l'integrale di f su $E \in \mathcal{A}$ è definito come

$$\int_E f(\omega) d\mu(\omega) = \sup_{\substack{s \text{ semplice} \\ 0 \leq s \leq f}} \int_E s(\omega) d\mu(\omega)$$

Se f misurabile qualsiasi allora, dati

$$f^+ = \max\{0, f\} \quad f^- = -\min\{0, f\},$$

$$\int_E f(\omega) d\mu(\omega) = \int_E f^+(\omega) d\mu(\omega) - \int_E f^-(\omega) d\mu(\omega) \quad (1)$$



Si definisce lo spazio delle funzioni integrabili rispetto ad una misura μ l'insieme

$$\mathcal{L}^1(\mu) = \{f: \Omega \rightarrow \mathbb{R} \text{ misurabile e con } \int_E |f(\omega)| d\mu(\omega) \leq +\infty\}$$



Sia I_i un intervallo di estremi a_i su \mathbb{R} e b_i con $a_i \leq b_i$. Si indica con $|I_i|$ la lunghezza in \mathbb{R} di I_i , ovvero $|I_i| = b_i - a_i$.

Sia $A = \prod_{i=1}^p I_i$ un intervallo di \mathbb{R}^n la cui lunghezza definita come

$$\lambda^n(A) = \prod_{i=1}^p |I_i|.$$

Se $B = \bigcup_{j=1}^q A_j$ è un pluriintervallo (gli A_j sono intervalli), la lunghezza di B è

$$\text{vol}(B) = \sum_{j=1}^q \lambda^n(A_j)$$



Per ogni sottoinsieme $A \in \mathbb{R}^n$ si può definire:

$$m^*(B) = \inf\{\text{vol}(M) : M \supseteq B\}$$

dove M è l'unione numerabile di prodotti di intervalli e $\text{vol}(M)$ è la somma dei prodotti delle lunghezze degli intervalli coinvolti. Si può dimostrare che m^* è una misura esterna. Si definisce A insieme misurabile secondo Lebesgue se

$$m^*(B) = m^*(A \cap B) + m^*(B \setminus A) \quad \forall B$$

Per il teorema di Carathéodory gli insiemi Lebesgue-misurabili formano una σ -algebra, e la misura di Lebesgue è definita da $m(A) = m^*(A)$ per ogni insieme Lebesgue-misurabile A .



La misura di probabilità P

$$P: \mathcal{A} \rightarrow [0, 1]$$

gode quindi delle seguenti proprietà

- ▶ $P(\emptyset) = 0$
- ▶ $P(\Omega) = 1$
- ▶ $A \in \mathcal{A} \implies P(A^c) = 1 - P(A)$
- ▶ $A, B \in \mathcal{A}, A \subseteq B \implies P(A) \leq P(B)$
- ▶ $A, B \in \mathcal{A} \implies P(A \cup B) = P(A) + P(B) - P(A \cap B)$

L'insieme \emptyset è detto **evento impossibile**.

L'insieme Ω è detto **evento certo**.

Un insieme misurabile A con $P(A) = 0$ si dice **evento trascurabile**.

Un insieme misurabile A con $P(A) = 1$ si dice **evento quasi certo**.



Considerando lo spazio di probabilità (Ω, \mathcal{A}, P) e lo spazio misurabile (Θ, \mathcal{B}) si dice che

- ▶ Ω è lo spazio campionario,
- ▶ $A \in \mathcal{A}$ è un evento,
- ▶ $\omega \in \Omega$ è un evento elementare,
- ▶ se A, B eventi e $A \cap B = \emptyset$ allora si parla di eventi incompatibili,
- ▶ $X: \Omega \rightarrow \Theta$ funzione misurabile è una variabile aleatoria. Se X dipende da qualche parametro si parla di processo stocastico. Se $\Theta = \mathbb{R}^n$ allora la variabile si dice n -dimensionale.



Data X variabile aleatoria

$$X: \Omega \rightarrow \Theta$$

si definisce **distribuzione/legge di probabilità** la misura di probabilità P_X come

$$P_X(V) = P(\{X \in V\}) = P(X^{-1}(V)) \quad \forall V \in \mathcal{B}$$

Se $\Theta = \mathbb{R}^n$ allora la legge di probabilità di X è individuata univocamente dalla sua **funzione di ripartizione** definita come

$$F(x) = P(X \leq \mathbf{x}) \quad \forall \mathbf{x} \in \mathbb{R}^n$$

Probabilità variabile aleatoria



Sia X la variabile aleatoria

$$X: \Omega \rightarrow \Theta$$

- X si dice discreta se Θ (il suo supporto o rango) può assumere un numero limitato di valori (al più numerabile). Si indica la *funzione di probabilità discreta* come

$$p(x) = P(\{X = \mathbf{x}\}) \quad \mathbf{x} \in \mathcal{B}$$

- X si dice continua se possiede supporto infinito. Si definisce la *funzione di densità di probabilità* f come

$$P(\{X \in A\}) = \int_A f(x) dx \quad A \in \mathcal{B}$$

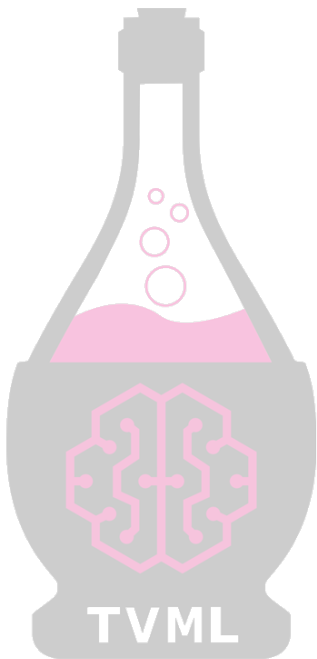


Dati A, B eventi, si definisce **probabilità condizionata**

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

I due eventi si dicono **indipendenti** se

$$P(A \cap B) = P(A) P(B)$$



Grazie dell'attenzione!