

Tutorato 03

Giulio Umbrella

Probabilita' condizionata

- Probabilita' condizionata
- Teorema probabilita' totale
- Chain rule
- Teorema di Bayes

Probabilita' condizionata

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

- Quindi stiamo considerando la probabilita' che due eventi accadano simultaneamente, normalizzando il valore in un intervallo $[0,1]$
- La probabilita' condizionata permette di calcolare una probabilita' in maniera molto piu' semplice

Chain rule

Dati due eventi A e B, possiamo scrivere $P(A, B) = P(A|B) \star P(B)$. Ma possiamo estendere questo ragionamento anche all'intersezione di piu' eventi, ad esempio:

$$P(A, B, C) = P(A) \star P(B|A) \star P(C|A, B)$$

Quindi stiamo trattando la probabilita' come una successione di eventi. Notare che in genere calcolare la probabilita' condizionata e' piu' **facile**, perche' abbiamo maggiori informazioni a disposizione.

Per una trattazione piu' formale, vedere la pagina wikipedia qui

Total probability theorem

Data una partizione dello spazio campionario tale che

1. $\Omega = B \cup B^c$
2. $B \cap B^c =$

Possiamo riscrivere la probabilita' di un evento A come

$$P(A) = P((A \cap B) \cup (A \cap B^c)) = P(A \cap B) + P(A \cap B^c) = P(A|B) \star P(B) + P(A|B^c) \star P(B^c)$$

Come con la chain rule, questa trattazione rende il calcolo della probabilita' piu' facile.

1. Spezziamo la probabilita' in sotto-casi
2. Utilizziamo la probabilita' condizionata per calcolare i modo piu' agevole i sotto-casi

Oss la chain-rule semplifica il calcolo di eventi congiunti, mentre il total probability theorem di eventi singoli.

Formula di Bayes

Esercizi

Esercizio 01

Un'urna contiene 6 palline rosse e 4 bianche. Vengono estratte **successivamente** due palline. Calcolare la probabilita' che siano entrambe rosse nell'ipotesi che ci sia o non ci sia reimmissione.

Sol

Stesso esercizio del tutorato 02 ma adesso abbiamo uno strumento in piu', la probabilita' condizionata. Possiamo definire un evento di interesse nel seguente modo:

$A = \{\text{Seconda pallina e' rossa dato che la prima pallina e' rossa}\}$

$$P(\text{Secondarossa} \cap \text{Primarossa}) = \frac{P(\text{Secondarossa}|\text{Primarossa})}{P(\text{Primarossa})}$$

- $P(\text{Prima rossa}) = \frac{6}{10}$
- $P(\text{Seconda rossa} | \text{Prima rossa}) = \frac{5}{9}$

La prima probabilita' e' immediata e deriva dall'approccio frequentistico. La seconda invece richiede un semplice ragionamento. Dato che sappiamo quale pallina e' stata estratta, conosciamo lo stato dell'urna alla seguente estrazione. Quindi utilizziamo questa informazione per calcolare la probabilita' di estrarre una seconda pallina rossa.

Esercizio 02

Si scelgono due carte a caso senza reimmissione da un mazzo di 52 carte. - D che entrambe le carte siano assi - P e' estratto l'asso di picche - A e' estratto almeno un asso

Determinare $P(D|P)$ e $P(D|A)$

Sol

Per $P(D|P)$ scriviamo la formula della probabilita' condizionata e ricaviamo numerato e denominatore.

$$P(D|P) =$$

Per $P(D|A)$ proseguiamo in modo analogo

$$P(D|A) =$$

Esercizio 03

Un'urna contiene 6 palline rosse e 4 bianche. Se ne estraggo 3 senza reimmissione. Qual'e' la probabilita' di ottenere B,N,N (in quest'ordine)?

Sol

Per risolvere questo esercizio usiamo la **chain rule**; la chain rule e' una semplice formula molto efficace per "smontare" un evento complesso nel prodotto di "sotto-eventi" la cui probabilita' e' piu' semplice da calcolare.

$$P(\text{Prima } B, \text{Seconda } N, \text{Terza } N) = P(\text{Prima } B) \star P(\text{Seconda } N | \text{Prima } B) \star P(\text{Terza } N | \text{Prima } B, \text{Seconda } N)$$

Ora calcoliamo le singole probabilita'; anche per questo esercizio usiamo le condizioni per rendere piu' semplici i calcoli in quanto ad ogni estrazione conosciamo lo **stato** dell'urna.

- $P(\text{Prima } B)$
- $P(\text{Seconda } N | \text{Prima } B)$
- $P(\text{Terza } N | \text{Prima } B, \text{Seconda } N)$

Esercizio 04

Consideriamo tre sacche contenenti ciascuna 100 palline:

- La sacca 1 ha 75 palline rosse e 25 blu
- La sacca 2 ha 60 palline rosse e 40 blu
- La sacca 3 ha 45 palline rosse e 55 blu

Scelgo una sacca a caso e poi estraggo una pallina. Qual'e' la probabilita' che sia rossa?

Sol

Posso considerare la probabilita' di estrarre la pallina rossa come la somma di tre casi distinti, a seconda della sacca che viene estratta. Quindi possiamo applicare il total probability theorem:

$$P(R) = P(R \cap S1) + P(R \cap S2) + P(R \cap S3)$$

E successivamente sostituire la probabilita' dell'intersezione usando la formula della probabilita' condizionata.

$$P(R) = P(R|S1) * P(S1) + P(R|S2) * P(S2) + P(R|S3) * P(S3)$$

Per ciascuna sacca ho la stessa probabilita' di scelta quindi $P(S1) = P(S2) = P(S3) = 1/3$

Una volta che ho l'informazione della sacca posso calcolare la probabilita' condizionata - $P(R|S1) = 75/100$ - $P(R|S2) = 60/100$ - $P(R|S3) = 45/100$

Infine sostituisco i valori nella formula di partenza

Oss Nel calcolo delle probabilita' condizionate $P(R|S1)$, $P(R|S2)$, $P(R|S3)$ **non** stiamo applicando meccanicamente la formule della probabilita' condizionata; stiamo infatti usando l'informazione per ricavare la probabilita' in modo **logico**.

Esercizio 05

Consideriamo tre sacche contenenti ciascuna 100 palline:

- La sacca 1 ha 75 palline rosse e 25 blu
- La sacca 2 ha 60 palline rosse e 40 blu
- La sacca 3 ha 45 palline rosse e 55 blu

Supponiamo che la pallina estratta sia la rossa. Qual'e' la probabilita' che la sacca scelta sia la 1? Cosa

Sol

Il testo dell'esercizio ci suggerisce che dobbiamo trovare la probabilita' condizionata $P(S1|R)$. Quindi usiamo la formula di Bayes per riscriverla in maniera piu' agevole

$$P(S1|R) = \frac{P(R|S1) * P(S1)}{P(R)}$$

Per maggiore chiarezza, riscriviamo la formula in modo completo:

$$P(S1|R) = \frac{P(R|S1) * P(S1)}{P(R|S1) * P(S1) + P(R|S2) * P(S2) + P(R|S3) * P(S3)}$$

Come prima cosa, vediamo che il valore al numeratore appare anche la denominatore; quindi il suo ruolo e' solo quello di **normalizzare** il valore fra 0 e 1.

Il valore di $P(A)$ e' lo stesso dell'esercizio 04, quindi e' gia' calcolato. Per il valore di $P(S1)$ invece sappiamo che e' lo stesso per ciascuna sacca, quindi e' pari a $1/3$

L'ultimo valore che ci manca e' $P(R|S1)$. Usiamo l'informazione che abbiamo sul numero della sacca e ricaviamo $P(R|S1) = 75/100$

$$P(S1|R) = \frac{0.75 * 1/3}{0.6} = 5/12 \approx 0.42$$

Commento

All'inizio dell'esperimento ciascuna sacca ha la stessa probabilita' e quindi $P(S1) = 1/3$. Ma abbiamo appena visto che la probabilita' di $P(S1|R)$ **aumenta** la probabilita' della sacca S1 a 0.4 circa.

1. Intuitivamente, la S1 e' quella con il maggior numero di pallina, quindi estrarre una pallina rossa ci
2. La $P(S1|R)$ e' la probabilita' ottenuta **dopo** aver ricevuto delle nuove informazioni. Quindi sulla base delle osservazioni empiriche, modifica il grado di confidenza che abbiamo nella probabilita' che la sacca sia la numero 1.

Come esercizio, calcolare le probabilita' $P(S2|R)$ e $P(S3|R)$. Quanto fa la somma delle tre probabilita'? Perche'?

Esercizio 06

Si consideri il seguente intrusion detection system (IDS) installato su un server. Un sistema ha due tipi di utente regolare e attaccante. Gli utenti si loggano nel sistema attraverso richieste da remoto. Lo scopo del sistema e' bloccare gli attaccanti e lasciare passare gli utenti organizzati.

La probabilita' che un attaccante sia riconosciuto dal sistema e' 0.99. La probabilita' che un utente autorizzato faccia scattare l'allarme e' 0.05. Una richiesta ogni 5000 e' un attacco.

Calcolare le seguenti probabilita'

1. Qual e' la probabilita' un utente sia bloccato.
2. Qual e' la probabilita' che una richiesta bloccata sia un attacco
3. Se una richiesta ha fatto scattare l'allarme, qual'e' la probabilita' che sia un attacco rispetto ad un momento prima che scattasse l'allarme?
4. Se nel server sono loggati 100 utenti e nessuno ha fatto scattare l'allarme, qual'e' la probabilita' che almeno uno sia un attaccante. Si supponga che gli attaccanti non coordinino le loro attivita'.

Sol

Riscriviamo le informazioni che abbiamo in modo probabilistico. In questo esercizio il passaggio fondamentale e' formalizzare in modo corretto le probabilita'.

Definiamo i seguenti eventi

- $A = \{\text{l'utente e' un attaccante}\}$
- $R = \{\text{l'utente e' regolare}\}$
- $B = \{\text{l'utente e' bloccato}\}$

La probabilita' che un attaccante sia bloccato dal sistema e' 0.99.

$$P(B|A) = 0.99$$

La probabilita' che un utente autorizzato sia bloccato e' 0.05

$$P(B|R) = 0.05$$

La probabilita' che l'utente sia un attaccante e'

$$P(A) = 1/5000 = 0.0002$$

Domanda 01 Qual'e' la probabilita' un utente sia bloccato.

La domanda chiede di calcolare $P(B)$. $P(B) = P(B \cap R) + P(B \cap A) = P(B|R) \star P(R) + P(B|A) \star P(A) = 0.05 \star 4999/5000 + 0.99 \star 1/5000 = 0.0502$

Abbiamo tutte le informazioni, tranne la probabilita' che l'utente sia regolare. Dato che abbiamo solo due tipi di utenti, possiamo ricavare questa probabilita' tramite evento **complementare**, $P(R) = 1 - P(A) = 1 - 0.0002$

Domanda 02 Qual'e' la probabilita' che un utente bloccato sia un attaccante

Di nuovo dobbiamo formalizzare la probabilita' nel modo seguente, $P(A|B)$. Per risolvere l'esercizio dobbiamo usare il teorema di Bayes

$$P(A|B) = \frac{P(B|A) \star P(A)}{P(B)} = \frac{0.99 \star 0.0002}{0.0502} = 0.0039$$

Domanda 03 Come cambia la probabilita'

La probabilita' che un utente sia un attaccante all'inizio e' $P(A) = 0.0002$; dopo che scatta la il blocco la probabilita' diventa $P(A|B) = 0.0039$. Quindi possiamo dire che un blocco aumenta il nostro grado di confidenza nel fatto che l'utente sia un attaccante.

Domanda 04

Oss la probabilita' che l'utente sia un attaccante dato un blocco e' 0.0039 un valore molto esiguo.