Tutorato 03

Giulio Umbrella

Argomenti di oggi

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

GOAL capire quale formulazione utilizzare in funzione del risultato che vogliamo ottenere

Intrusion detection system 01

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

Intrusion detection system 02

Performance del sistema

- Il sistema identifica e blocca un attaccante nel 99 % dei casi
- Il 99.61 % degli utenti bloccati sono regolari

Sondaggio

1. Il sistema funziona bene
2. Il sistema funziona male
3. Non mi interessa la cybersecurity

Probabilita' condizionata

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

Due osservazioni

1. P(B) > 0
2. Stiamo ipotizzando che l'evento B accada in una sorta di **esperimento mentale**

Probabilita' condizionata - Intuizione 01

- **Domanda** Dato un dado a sei facce, qual'e' la probabilita' che venga estratto un numero maggiore o uguale a quattro, sapendo che il numero estratto e' pari?
- Definiamo uno **spazio campionario** e usiamo i **diagrammi** per motivare la risposta

Probabilita' condizionata - Intuizione 02

Riprendiamo due concetti gia' visti

Intersezione fra insiemi

L'intersezione fra due insiemi $A \cap B$ non puo' essere piu' grande degli insiemi di partenza. Quali sono i possibili casi?

Normalizzazione

La normalizzazione serve per riportare i valori fra 0 e 1

Probabilita' condizionata - Intuizione 03

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

- 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

Probabilita' condizionata - come si usa

Usando direttamente la formula

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

Riscrivendo la formula

- $P(A \cap B) = P(A|B) \star P(B)$

Probabilita' condizionata - dove si usa

Con il teorema della probabilita' totale

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

Con la chain rule

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

Teorema di Bayes

- $P(H|D) = \frac{P(D|H) \star P(H)}{P(D)}$

Esercizio 01

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)

Esercizio 02

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.

Chain rule

- $P(A, B, C) = P(A) \star P(B|A) \star P(C|A, B)$
- Proviamo a derivarla in maniera intuitiva; il trucco e' pensare a $B \cap C$ come ad un singolo evento X.

Osservazioni

1. Puo' essere applicata per N eventi
2. Comoda quando abbiamo una **sequenza** di eventi

Esercizio 03

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

Teorema della probabilita' totale

- $\mathrm{P(A)} = P((A \cap B) \cup P(A \cap B^c)) = P(A \cap B) + P(A \cap B^c) = P(A|B) \star P(B) + P(A|B^c) \star P(B^c)$

Osservazioni

1. Usiamo questa formula quando vogliamo **spezzettare** la probabilita' in sotto-casi mutualmente esclusivi

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. Quale' la probabilita' che sia rossa?

Esercizio 04 Commento

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**.

Teorema di Bayes

$$P(H|D) = \frac{P(D|H) \star P(H)}{P(D)}$$

- Il denominatore serve a **normalizzare** il valore fra 0 e 1, ma non ha valore informativo
- La probabilita' condizionata dipende da due fattori la probabilita' dell'ipotesi e la probabilita' di osservare il dato se l'ipotesi e' verificata.

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, **senza** farvi vedere la sacca dalla quale la pallina e' estratta. Qual'e' la probabilita' che la sacca scelta sia la 1?

Esercizio 05 Commento 01

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.

Esercizio 05 Commento

Cosa possiamo dire delle altre tre sacche?

- $P(S2|R) = \frac{P(R|S2) \star P(S2)}{P(R)} = \frac{0.6 \star 1/3}{0.6} = 0.33$
- $P(S3|R) = \frac{P(R|S3) \star P(S3)}{P(R)} = \frac{0.45 \star 1/3}{0.6} = 0.25$

Commento

- La somma delle tre probabilita' da come risultato 1.
- **IMPORTANTE**: se condizioniamo su evento, otteniamo una distribuzione di probabilita' a tutti gli effetti

Esercizio 06

Si consideri il seguente intrusion detection system (IDS) installato su un server. Un sistema ha due tipi di utente **autorizzato** 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 regolare.

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

Calcolare le seguenti probabilita'

1. Qual'e' la probabilita' che un utente sia bloccato
2. Qual'e' la probabilita' che un utente bloccata sia un attaccante
3. Se un utente viene bloccato, qual'e' la probabilita' che sia un attaccante rispetto ad un momento prima in cui scattasse il blocco?
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 coordino le loro

Esercizio 06 punto 01

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

Esercizio 06 punto 02

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

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

Esercizio 06 punto 03

La probabilita' che un utente sia un attaccante all'inizio e' $P(A) = 0.0002$; dopo che scatta la il blocco la probabilita' diventa $P(A|B) = 0.0039$

Esercizio 06 Commento

Abbiamo due valori che sembrano essere in contraddizione

1. P(Blocco utente |Utente e' Attaccante) = 0.99
2. P(Utente e' Attaccante| Blocco utente) = 0.0039

Sondaggio

1. Il sistema funziona bene
2. Il sistema funziona male
3. Non mi interessa la cybersecurity

Esercizio 06 Commento

- Quindi, il sistema riesce ad individuare gli utenti molto bene. Ma al tempo stesso la maggior parte dei blocchi sono **falsi positivi**, perche' la probabilita' che un utente bloccato sia un attaccante e' molto bassa.
- Il problema e' che la probabilita' che un utente sia un attaccante e' molto bassa! Talmente basse che il sistema non e' abbastanza affidabile.

Quindi, il sistema riesce a riconoscere e bloccare gli attaccanti. Ma gli attaccanti sono talmente pochi che la maggior parte dei blocchi va a colpire utenti regolari.