Lezione 2

Probabilità Condizionale e Teorema di Bayes

Fondamenti di Calcolo delle Probabilità

Corso	Unità
Data Analyst e AI per Programmatori	IA.2

AGGIORNARE LE PROBABILITÀ CON NUOVE INFORMAZIONI

Avete mai sentito parlare di uno yacht chiamato "Bayesian"?

E sapete perché porta questo nome così particolare, profondamente legato al mondo delle probabilità?

Preparatevi a scoprire una storia affascinante che collega il lusso, la logica e una delle menti più influenti della statistica.

Mike Lynch: un Capitano della Probabilità

Dietro il nome **"Bayesian"** si cela la mente brillante di **Mike Lynch**, un imprenditore e visionario britannico protagonista in passato del mondo del software e dell'intelligenza artificiale.

Per Lynch, il Teorema di Bayes non era un semplice esercizio accademico. Dopo averlo esplorato a fondo per la sua tesi di dottorato in matematica presso l'Università di Cambridge, divenne la pietra angolare della sua filosofia aziendale.

> "Ho chiamato lo yacht 'Bayesian' perché il Teorema di Bayes era alla base di tutto ciò che facevo, dalla mia tesi alla tecnologia di Autonomy. Era un modo per celebrare l'eleganza e la potenza di questa idea."

La sua azienda, **Autonomy Corporation**, ha rivoluzionato il modo in cui le macchine comprendevano i dati non strutturati. Utilizzando complessi algoritmi bayesiani per analizzare e interpretare il "significato" di testi, email e conversazioni, Autonomy è diventata un colosso tecnologico, dimostrando come la probabilità condizionale possa essere una forza motrice per l'innovazione e il successo.

Aggiornare le Probabilità con Nuove Informazioni

La **probabilità condizionale** è uno degli strumenti più potenti della statistica moderna. Risponde a una domanda fondamentale che emerge continuamente nella vita quotidiana, nella scienza e nell'intelligenza artificiale:

> **"Come cambia la probabilità di un evento quando otteniamo nuove informazioni?"**

Questa capacità di aggiornare le nostre credenze in base alle evidenze è alla base del ragionamento razionale e del metodo scientifico.

📊 Esempio Motivante: Previsioni Meteo

Situazione iniziale: In una certa città italiana, il 23% dei giorni è piovoso. Quindi **P(Piove) = 0.23**

Nuova informazione: Oggi il cielo è nuvoloso e grigio. Come cambia la nostra previsione?

Domanda: Qual è ora **P(Piove | Nuvoloso) = ?**

Intuitivamente sappiamo che la probabilità di pioggia deve aumentare, ma di quanto? La probabilità condizionale ci fornisce gli strumenti matematici per rispondere con precisione.

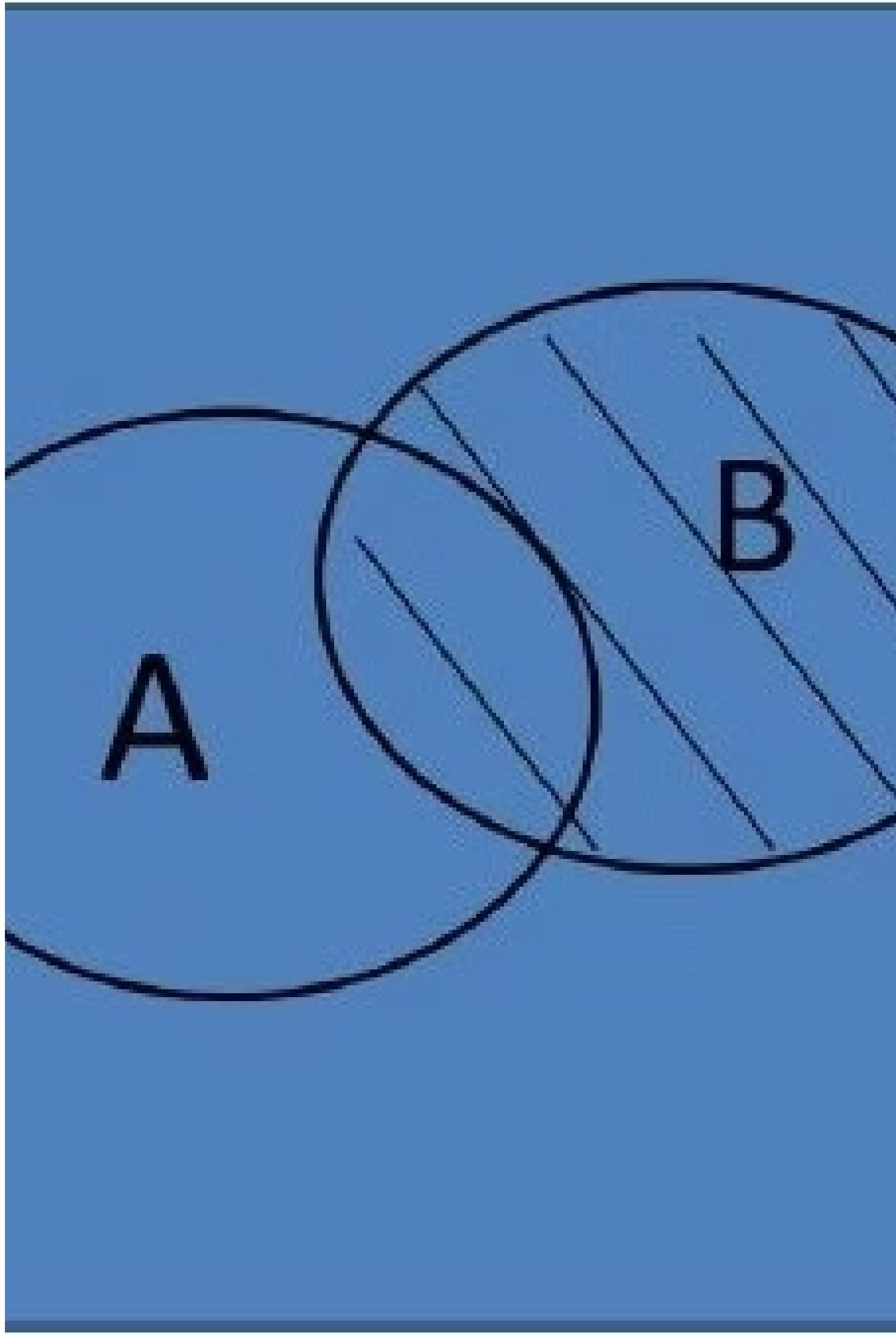
Eventi Condizionati

Quando parliamo di eventi condizionati, intendiamo situazioni in cui il verificarsi di un evento può influenzare la probabilità che se ne verifichi un altro.

Ad esempio, se abbiamo due eventi A e B, diciamo che B condiziona A — e lo scriviamo come $(A|B)$ — quando sapere che B è accaduto cambia la probabilità che accada A.

Si legge **"A dato B"**: in questo caso A è l'evento condizionato, mentre B è l'evento condizionante.

Perché il condizionamento abbia senso, B deve avvenire prima di A, quindi esiste un ordine temporale che va da B ad A.



Definizione di Probabilità Condizionale

Definizione Formale

Dati due eventi **A** e **B** in uno spazio campionario **S**, la **probabilità condizionale di A dato B** è definita come:

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

quando $P(B) > 0$

Notazione

P(A|B) si legge in due modi equivalenti:

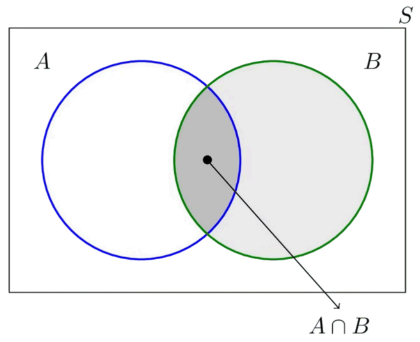
- "probabilità di A **dato** B"
- "probabilità di A **condizionata** a B"

Questa notazione è universale in matematica, statistica e machine learning.

Terminologia Chiave

- **P(B)** = probabilità a priori (prior)
- **P(A|B)** = probabilità a posteriori (posterior)

Intuizione Geometrica



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

💡 Spiegazione Intuitiva

Quando sappiamo che **B** è avvenuto, lo spazio campionario si **riduce** all'insieme **B**.

Ora **A** può avvenire solo se l'esito appartiene all'intersezione **A ∩ B**.

Dividiamo per **P(B)** per "rinormalizzare" lo spazio ridotto, garantendo che $P(B|B) = 1$.

La probabilità condizionale è il fondamento matematico per l'aggiornamento razionale delle credenze basato su nuove evidenze.

Esempio: Lancio di un Dado

Vediamo come applicare la definizione di probabilità condizionale a un esempio concreto e intuitivo.

📊 Problema

Lancio un dado equo a sei facce. Definisco due eventi:

A = {1, 3, 5} → numeri **dispari**

B = {1, 2, 3} → numeri **≤ 3**

Domanda

Qual è **P(A|B)**?

🎯 Interpretazione

Sapendo che il risultato del lancio è **≤ 3**, qual è la probabilità che sia anche **dispari**?

In altre parole: se il dado mostra 1, 2 o 3, quante probabilità ci sono che sia 1 o 3 (cioè dispari)?

Questo problema ci permette di vedere come l'informazione aggiuntiva (il risultato è ≤ 3) modifica la probabilità originale che il numero sia dispari.

Esempio: Lancio di un Dado - Soluzione

01

Probabilità a priori di A

Senza alcuna informazione aggiuntiva, la probabilità di ottenere un numero dispari è:

$P(A) = 3/6 = 1/2$

Tre esiti favorevoli (1, 3, 5) su sei possibili.

02

Calcolo dell'intersezione A ∩ B

Quali sono i numeri che sono sia dispari che ≤ 3?

$A \cap B = \{1, 3\}$

$P(A \cap B) = 2/6 = 1/3$

03

Probabilità di B

La probabilità che il risultato sia ≤ 3 è:

$P(B) = 3/6 = 1/2$

Tre esiti favorevoli (1, 2, 3) su sei possibili.

04

Applicazione della formula

Ora applichiamo la definizione di probabilità condizionale:

$P(A|B) = P(A \cap B) / P(B) = (1/3) / (1/2) = 2/3$

Risultato e Interpretazione

Sapendo che il risultato è ≤ 3, la probabilità che sia dispari **aumenta** da 1/2 a 2/3!

Questo ha senso: nello spazio ridotto {1, 2, 3}, due elementi su tre sono dispari (1 e 3), mentre solo uno è pari (2).

Esempio: Lancio di Due Dadi

Aumentiamo la complessità con un lancio di due dadi, per applicare ulteriormente il concetto di probabilità condizionale.

📊 Problema

Lanciamo una coppia di dadi equa a sei facce. Definiamo due eventi:

A = "almeno uno dei due dadi mostra un **2**"

B = "la **somma** dei due dadi è **6**"

Domanda

Qual è **P(A|B)**?

🎯 Interpretazione

Sapendo che la somma dei due dadi è esattamente **6**, qual è la probabilità che almeno uno dei due dadi abbia dato come esito un **2**?

In altre parole: se i risultati possibili sono solo quelli la cui somma è 6, quanti di questi includono un 2?

Questo problema ci aiuta a capire come l'informazione "la somma è 6" restringe lo spazio campionario e influenza la probabilità che si verifichi l'evento A.

Esempio: Lancio di Due Dadi - Soluzione

01

Spazio Campionario Ridotto (B)

L'evento **B** ("la somma dei due dadi è 6") riduce il nostro spazio campionario a 5 esiti favorevoli tra i 36 totali:

- (1, 5)
- (2, 4)
- (3, 3)
- (4, 2)
- (5, 1)

Quindi, la probabilità di B è **P(B) = 5/36**.

02

Intersezione A ∩ B

Ora identifichiamo gli esiti che soddisfano sia **B** che **A** ("almeno uno dei due dadi mostra un 2"). Tra gli esiti della somma 6, quali contengono un 2?

- (2, 4)
- (4, 2)

Ci sono **2** esiti favorevoli per l'intersezione. Pertanto, **P(A ∩ B) = 2/36**.

03

Applicazione della Formula

Applichiamo la formula della probabilità condizionale:

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

Sostituendo i valori calcolati:

$$P(A|B) = \frac{2/36}{5/36} = \frac{2}{5}$$

Risultato e Interpretazione

Sapendo che la somma dei due dadi è 6, la probabilità che almeno uno dei dadi mostri un 2 è **2/5** (o 40%).

Questo dimostra come l'informazione aggiuntiva sulla somma modifichi significativamente la probabilità di osservare un 2.

Verifica la tua Comprensione

È il momento di mettere alla prova le tue conoscenze sulla probabilità condizionale con alcune domande e un breve problema pratico.

Domande Concettuali

1. Qual è la formula per calcolare la probabilità condizionale $P(A|B)$?
2. Spiega la differenza tra probabilità "a priori" (prior) e "a posteriori" (posterior) nel contesto della formula $P(A|B)$.
3. In quale scenario $P(A|B) = P(A)$? Cosa implica questo per gli eventi A e B?

Problema Pratico: Palline nel Sacchetto

Immagina di avere un sacchetto con 10 palline:

- 6 rosse (R) e 4 blu (B).
- Di queste, 3 palline rosse sono grandi (G) e 3 sono piccole (P).
- Tutte le palline blu sono piccole (P).

Se estrai una pallina dal sacchetto e sai che è piccola (P), qual è la probabilità che sia rossa (R)?

Calcola **$P(R|P)$**.

Verifica la tua Comprensione: Risposte Teoriche

1. Formula della Probabilità Condizionale

La formula per calcolare la probabilità condizionale P(A|B) è:

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

Questo vale a condizione che $P(B) > 0$. Rappresenta la probabilità che A si verifichi, dato che B si è già verificato.

2. Probabilità "a Priori" vs "a Posteriori"

- **A Priori (Prior)**: La probabilità iniziale di un evento o di una condizione (es. $P(B)$) *prima* di considerare nuove informazioni o evidenze.
- **A Posteriori (Posterior)**: La probabilità aggiornata di un evento (es. $P(A|B)$) *dopo* aver tenuto conto di nuove informazioni o evidenze. Riflette la nostra credenza modificata.

3. Quando P(A|B) = P(A)?

Se $P(A|B) = P(A)$, significa che gli eventi **A** e **B** sono **indipendenti**.

L'indipendenza implica che il verificarsi dell'evento B non modifica in alcun modo la probabilità di A. In questo caso, $P(A \cap B) = P(A) \cdot P(B)$.

Queste risposte consolidano i concetti fondamentali che abbiamo esplorato, gettando le basi per affrontare problemi più complessi.

Esempio: Palline nel Sacchetto - Soluzione

01

1. Identifica la probabilità di estrarre una pallina piccola (P)

Il sacchetto contiene 10 palline in totale.

Le palline piccole sono:

- 3 rosse piccole
- 4 blu piccole

Quindi, ci sono 7 palline piccole. La probabilità di estrarre una pallina piccola è:

$$P(P) = \frac{7}{10}$$

02

2. Identifica la probabilità dell'intersezione (R ∩ P)

Vogliamo trovare le palline che sono sia rosse che piccole. Dal problema sappiamo che:

- 3 palline rosse sono piccole

La probabilità di estrarre una pallina rossa e piccola è:

$$P(R \cap P) = \frac{3}{10}$$

03

3. Applica la formula della probabilità condizionale

La formula per la probabilità condizionale P(R|P) è:

$$P(R|P) = \frac{P(R \cap P)}{P(P)}$$

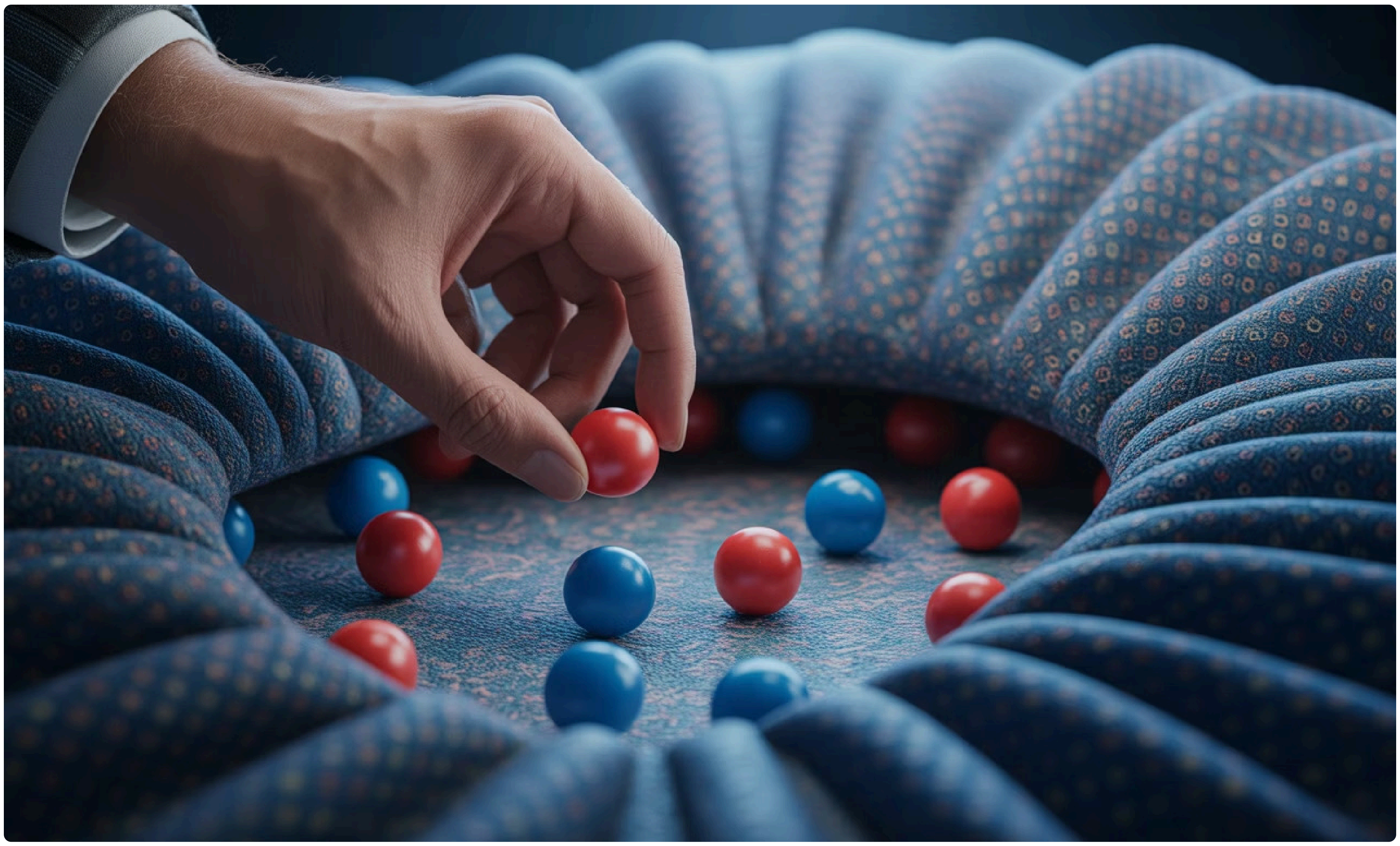
Sostituendo i valori calcolati:

$$P(R|P) = \frac{3/10}{7/10} = \frac{3}{7}$$

Risultato e Interpretazione

Sapendo che la pallina estratta è piccola, la probabilità che sia rossa è **3/7** (circa 42.86%).

Questo dimostra come l'informazione aggiuntiva ("la pallina è piccola") riduce lo spazio campionario da 10 a 7 palline, e all'interno di questo nuovo spazio, 3 sono rosse.



Esempio di Codice: Calcolo P(A|B)

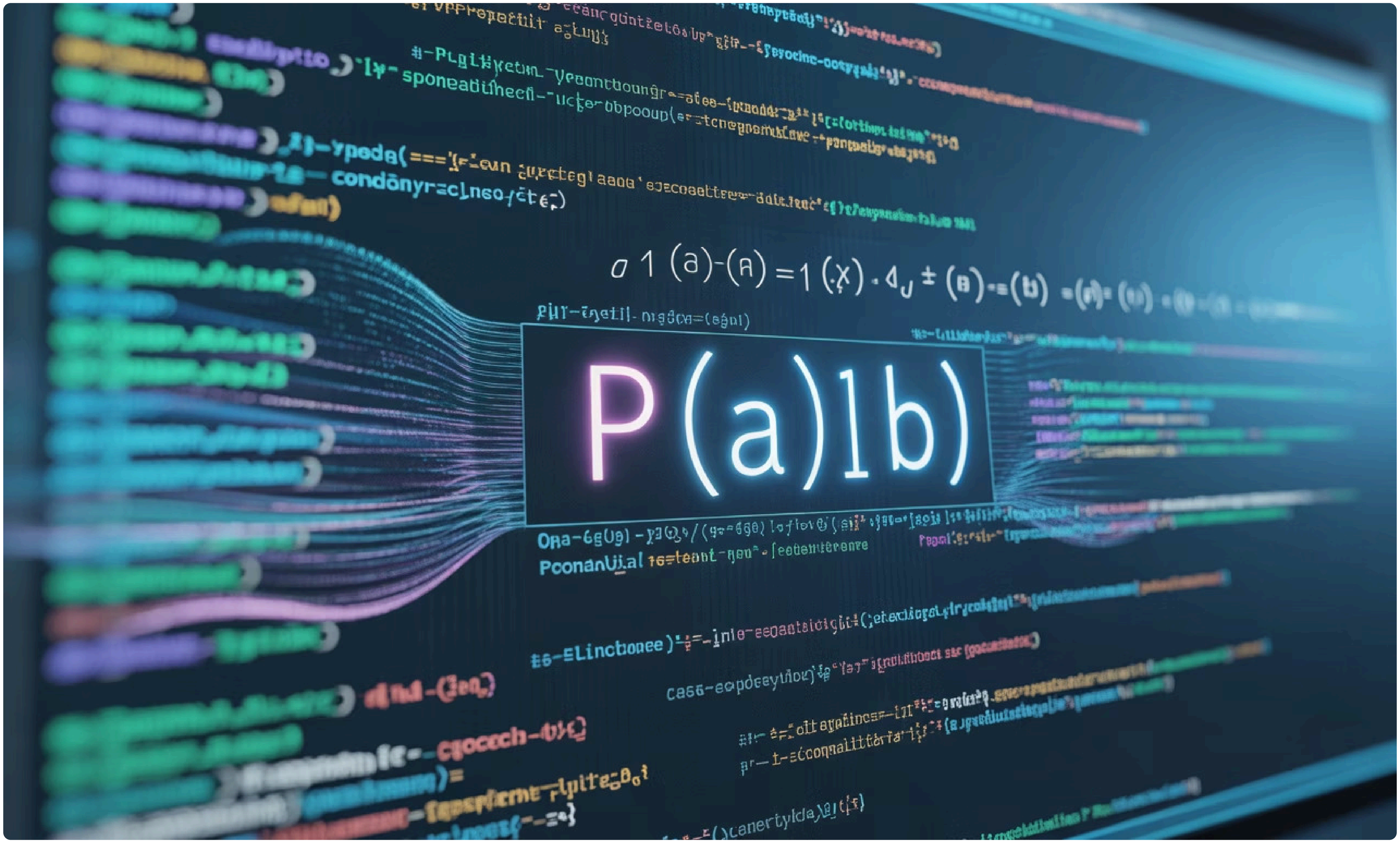
Passiamo ora a un esempio pratico di come calcolare la probabilità condizionale P(A|B) tramite codice. Questo programma è stato scritto per automatizzare il calcolo una volta fornite le probabilità necessarie, offrendo un ponte tra la teoria e l'applicazione pratica.

Esercizio 1 Verifica dei Risultati

Verificare che il risultato calcolato automaticamente sia coerente con quello ottenuto tramite il calcolo manuale, consolidando la comprensione dei principi sottostanti.

Eserczio 2 Inserisci Input da codice delle Probabilità

Modifica il codice ed inserisci direttamente P(A∩B) e P(B) come input, eliminando la necessità di inserimenti manuali ripetuti durante l'esecuzione.



Esercizio 1: modifica il codice sulla probabilità condizionata per lo scenario Auto a Noleggio

Applichiamo la probabilità condizionale con questo scenario pratico. Una compagnia di autonoleggio offre una flotta diversificata di veicoli:

Dettaglio Flotta

- **Totale Auto:** 100 veicoli
- **City Car:**
  - 50 Toyota
  - 20 Audi

Dettaglio SUV

- **SUV:**
  - 25 Toyota
  - 5 Audi

Considerando questa flotta, qual è la probabilità che l'auto fornita sia un SUV, se sappiamo che è un'Audi?

Calcola **P(SUV|Audi)**.

Esercizio 2: modifica il codice sulla probabilità condizionata per lo scenario Ristorante

Applichiamo la probabilità condizionale a uno scenario di ordini in un ristorante. Dati i seguenti 300 ordini, vogliamo determinare la probabilità che un ordine sia vegetariano, sapendo che è stato richiesto per l'asporto.

Veg	90	54
Carne	160	56
Pesce	50	20
Totale	300	130

Calcola la probabilità condizionale: **P(Veg | Takeaway)**

Regola del Prodotto (Chain Rule)

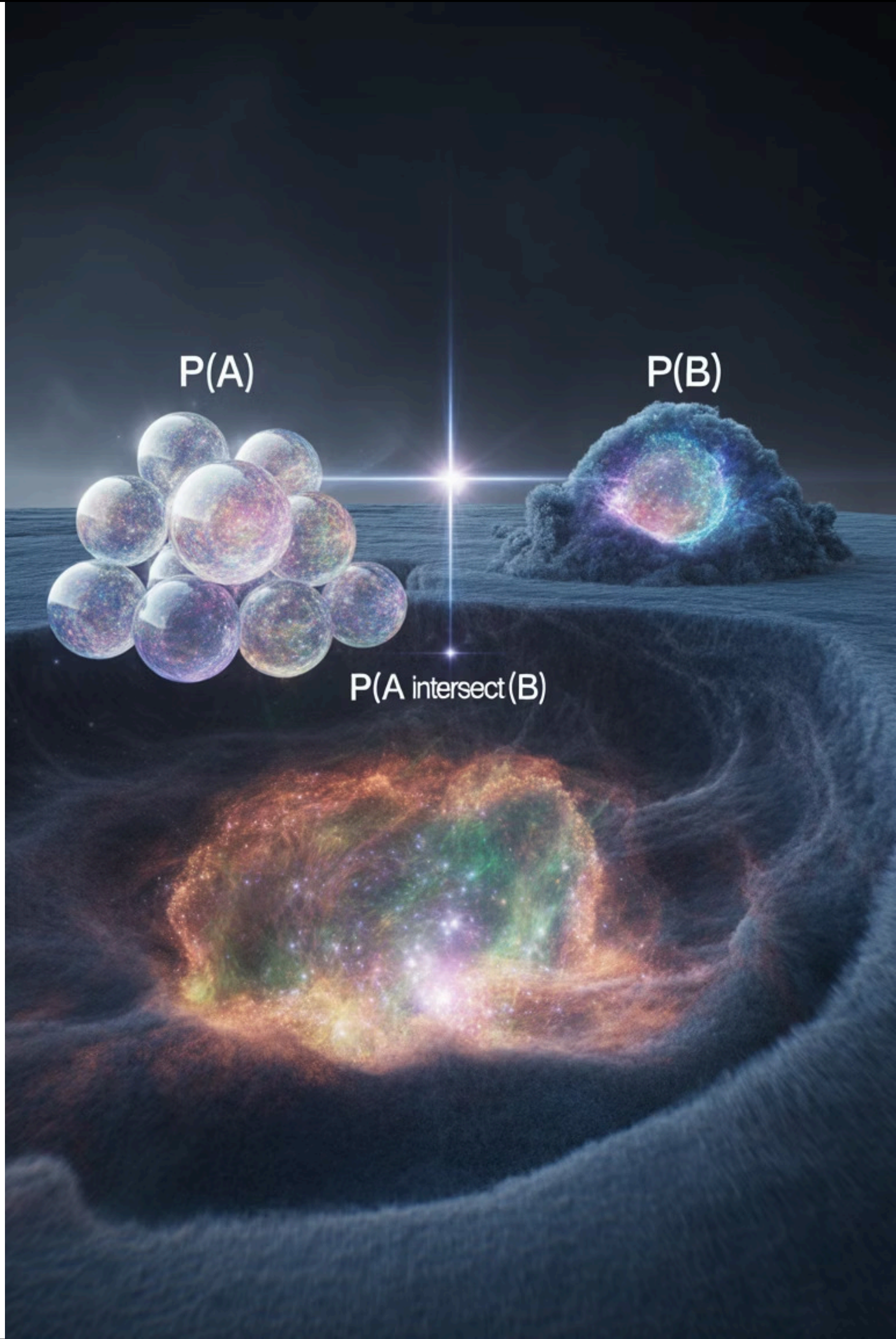
La regola del prodotto (o delle probabilità composte) è fondamentale per calcolare la probabilità che due eventi, **A** e **B**, si verifichino contemporaneamente.

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

Oppure, in modo equivalente:

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

Questa formula esprime che la probabilità congiunta dei due eventi è data dalla probabilità del primo evento, moltiplicata per la probabilità del secondo, sapendo che il primo è già avvenuto. È detta anche "regola delle probabilità composte" perché combina più eventi in un'unica misura di probabilità.



Regola del Prodotto (Chain Rule)

Dalla definizione di probabilità condizionale possiamo derivare una formula molto utile per calcolare la probabilità dell'intersezione di eventi.

Estensione a Tre Eventi

La regola si estende naturalmente a sequenze più lunghe:

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B)$$

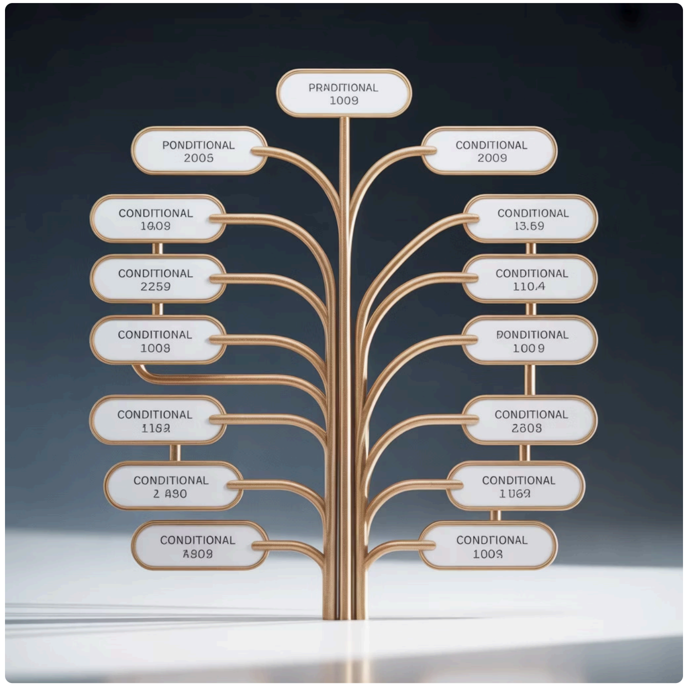
dove $P(C|A \cap B)$ è la probabilità di C dato che sia A che B sono avvenuti.

💡 Utilità Pratica

Spesso è molto più facile calcolare le probabilità condizionali **P(A|B)** e **P(B)** separatamente che calcolare direttamente **P(A ∩ B)**.

Questa formula ci permette di scomporre problemi complessi in parti più semplici e intuitive.

Visualizzazione: Diagramma ad Albero



Come usare il diagramma:

Per ottenere la probabilità di un percorso completo nell'albero, **moltiplichiamo** le probabilità lungo tutti i rami del percorso.

Esempio: $P(B \cap A) = P(B) \times P(A|B)$

I diagrammi ad albero sono particolarmente utili per visualizzare sequenze di eventi dipendenti e calcolare probabilità complesse in modo sistematico.

Esempio Pratico: Estrazione da un'Urna con Diagramma ad Albero

Per comprendere meglio l'applicazione della Regola del Prodotto e come i diagrammi ad albero possono semplificare calcoli complessi, analizziamo un caso concreto.

01

Scenario: Estrazione di Due Palline

Immagina di avere un'urna contenente un totale di **5 palline**:

- 3 palline rosse (R)
- 2 palline blu (B)

Estraiamo due palline una dopo l'altra, **senza rimettere la prima pallina** nell'urna. Il nostro obiettivo è calcolare la probabilità che **entrambe le palline estratte siano rosse** ($P(R_1 \cap R_2)$).

02

Costruzione del Diagramma ad Albero

Il diagramma ci aiuta a visualizzare le probabilità a ogni passo:

1° estrazione:

- $P(R_1) = 3/5$ (3 rosse su 5 totali)
- $P(B_1) = 2/5$ (2 blu su 5 totali)

2° estrazione (condizionale):

- Se la 1° era R: $P(R_2|R_1) = 2/4$ (restano 2 rosse su 4 totali)
- Se la 1° era B: $P(R_2|B_1) = 3/4$ (restano 3 rosse su 4 totali)

03

Calcolo della Probabilità Composta

Per trovare la probabilità che entrambe le palline siano rosse ($R_1 \cap R_2$), seguiamo il percorso "R" e poi "R" sul diagramma ad albero e moltiplichiamo le probabilità lungo i rami:

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2|R_1)$$

Sostituendo i valori derivati dalla costruzione del diagramma:

$$P(R_1 \cap R_2) = \frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20} = \frac{3}{10} = 0.3$$

Conclusione

La probabilità di estrarre due palline rosse consecutivamente da quest'urna, senza reimmissione, è del **30%**.

Questo esempio illustra chiaramente come la probabilità di un evento successivo sia influenzata dall'esito dell'evento precedente (probabilità condizionale).

Legge della Probabilità Totale

Quando vogliamo calcolare la probabilità di un evento complesso, possiamo scomporlo considerando tutti i possibili "scenari" mutuamente esclusivi attraverso cui può verificarsi.

Partizione dello Spazio Campionario

Gli eventi **B_1, B_2, ..., B_n** formano una **partizione** di S se:

1. **Sono disgiunti:** $B_i \cap B_j = \emptyset$ per $i \neq j$
2. **Coprono S:** $B_1 \cup B_2 \cup ... \cup B_n = S$

Formula

$$P(A) = \sum_i P(A|B_i) \cdot P(B_i)$$

Espandendo esplicitamente:

$$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + ... + P(A|B_n) \cdot P(B_n)$$

💡 Intuizione

L'evento **A** può avvenire in **modi mutuamente esclusivi** corrispondenti ai **B_i**. Sommiamo le probabilità pesate per ciascun modo possibile.

Questa formula è essenziale per applicare il teorema di Bayes quando dobbiamo calcolare **P(A)** al denominatore.

📊 Esempio: Difetti di Produzione

Un'azienda manifatturiera ha tre fabbriche che producono lo stesso componente:

- **Fabbrica 1:** produce il 30% dei componenti totali, con 1% di prodotti difettosi
- **Fabbrica 2:** produce il 45% dei componenti totali, con 2% di prodotti difettosi
- **Fabbrica 3:** produce il 25% dei componenti totali, con 3% di prodotti difettosi

Domanda

Se prendiamo un prodotto a caso dalla produzione totale, qual è la probabilità che sia difettoso?

P(Difettoso) = ?

Soluzione

Applichiamo la legge della probabilità totale:

$P(D) = P(D|F_1) \cdot P(F_1) + P(D|F_2) \cdot P(F_2) + P(D|F_3) \cdot P(F_3)$

$= 0.01 \times 0.30 + 0.02 \times 0.45 + 0.03 \times 0.25$

$= 0.003 + 0.009 + 0.0075$

= 0.0195 = 1.95%

Quindi circa il 2% dei prodotti totali è difettoso.

Teorema di Bayes: Invertire la Condizionalità

Il teorema di Bayes è uno dei risultati più importanti e applicati della teoria delle probabilità. Ci permette di "invertire" le probabilità condizionali, passando da **P(A|B)** a **P(B|A)**.

$$P(B|A) = [P(A|B) \cdot P(B)] / P(A)$$

P(B) = Prior (A Priori)

La nostra conoscenza o credenza iniziale su B **prima** di osservare l'evidenza A. Rappresenta ciò che sappiamo indipendentemente dai dati.

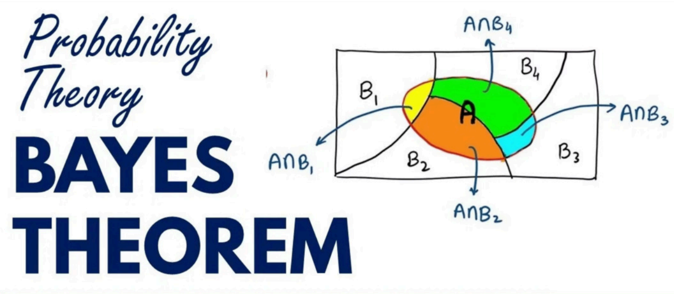
P(A|B) = Likelihood (Verosimiglianza)

Quanto sono **compatibili** i dati osservati A con l'ipotesi B? Quanto è probabile osservare A se B fosse vero?

P(B|A) = Posterior (A Posteriori)

La nostra conoscenza **aggiornata** su B dopo aver osservato l'evidenza A. Il risultato finale che cerchiamo.

🎯 Paradigma Bayesiano



Il teorema di Bayes formalizza matematicamente il processo di apprendimento e aggiornamento razionale delle credenze.

Applicazioni Fondamentali

- **Medicina:** Diagnosi mediche e interpretazione di test
- **Machine Learning:** Classificatori bayesiani e inferenza
- **Finanza:** Valutazione del rischio e decisioni di investimento
- **Giustizia:** Valutazione di prove ed evidenze forensi
- **Scienza:** Verifica di ipotesi e aggiornamento di teorie

Teorema di Bayes: Il Navigatore della Conoscenza

Il teorema di Bayes è come un **navigatore delle probabilità**: parte da dove pensavi di essere, riceve un nuovo segnale (nuove informazioni o evidenze), e ti dice in modo razionale come cambiare rotta nella tua convinzione.

Bayes ci insegna a **cambiare idea in modo razionale** quando arrivano nuove informazioni, aggiornando le nostre probabilità iniziali in base all'evidenza.

$$P(Causa|Evidenza) = \frac{P(Evidenza|Causa) \cdot P(Causa)}{P(Evidenza)}$$

➡️ Questa formula ci permette di aggiornare la probabilità di una causa (o ipotesi) dopo aver osservato un determinato effetto (o evidenza), rendendo la nostra comprensione più precisa e informata.

Teorema di Bayes: Tre Metafore Chiave

Per comprendere meglio l'intuizione alla base del Teorema di Bayes, esploriamo tre metafore pratiche che illustrano il suo potere di aggiornare le nostre credenze in base a nuove evidenze.

Il Detective

Hai un sospetto iniziale (probabilità a priori). Trovi un'impronta sulla scena del crimine (nuova evidenza). Bayes quantifica di quanto aumenta la probabilità che il sospetto sia il colpevole, aggiornando la tua credenza in base a questa prova concreta.

Aggiornamento della credenza sulla base di una prova.

Il Medico

Solo 1 persona su 1000 ha una rara malattia. Il test di screening risulta positivo (evidenza). Bayes ti aiuta a rispondere: "Sapendo che il test è positivo, qual è la probabilità reale di avere la malattia?", correggendo l'intuizione che spesso ignora i falsi positivi.

Corregge l'intuizione tenendo conto dei falsi positivi.

Il Meteo

Storicamente piove 1 giorno su 10. Oggi vedi il cielo completamente coperto (nuova evidenza). Bayes ti permette di calcolare come la probabilità di pioggia passi dal 10% iniziale a una previsione aggiornata, basandosi sull'osservazione attuale del cielo.

Aggiorna la previsione in base a ciò che osservi.

False Positive Paradox: Test Diagnostico

Uno degli esempi più sorprendenti e controintuitivi dell'applicazione del teorema di Bayes è il **paradosso dei falsi positivi**. Questo caso dimostra come anche test molto accurati possano dare risultati fuorvianti quando la condizione testata è rara.

📊 Scenario Clinico

Consideriamo una malattia rara che colpisce **1 persona su 10.000** nella popolazione generale. È stato sviluppato un test diagnostico molto accurato con le seguenti caratteristiche tecniche:

Prevalenza

P(Malato) = 0.0001

Solo 1 persona su 10.000

Questa è la probabilità a priori di avere la malattia nella popolazione generale.

Sensibilità

P(Positivo | Malato) = 0.99

Accuratezza del 99%

Il test identifica correttamente il 99% delle persone effettivamente malate (veri positivi).

Specificità

P(Negativo | Sano) = 0.98

Quindi **P(Positivo | Sano) = 0.02**

Il test è negativo nel 98% delle persone sane, ma dà un falso positivo nel 2% dei casi.

❓ Domanda Chiave

Una persona viene scelta a caso dalla popolazione e risulta **positiva** al test diagnostico.

Qual è la probabilità che sia effettivamente malata?

In altre parole: **P(Malato | Positivo) = ?**

Pensa a quale potrebbe essere la risposta prima di vedere la soluzione nella prossima slide....

False Positive Paradox: Soluzione e Spiegazione

💡 Calcolo con Teorema di Bayes

Applichiamo il teorema di Bayes usando M = Malato e T = Test Positivo:

$$P(M|T) = \frac{P(T|M) \cdot P(M)}{P(T|M) \cdot P(M) + P(T|M^c) \cdot P(M^c)}$$

Sostituiamo i valori numerici:

= (0.99 × 0.0001) / [(0.99 × 0.0001) + (0.02 × 0.9999)]

= 0.000099 / [0.000099 + 0.019998]

= 0.000099 / 0.020097

≈ 0.0049

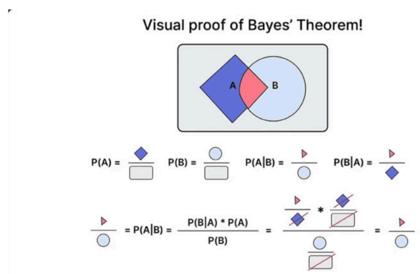
P(Malato | Positivo) ≈ 0.49%

⚠️ Risultato Sorprendente!

Anche con un test accurato al **98-99%**, la probabilità di essere effettivamente malato dato un test positivo è solo dello **0.49%**!

Questo significa che **circa 199 persone su 200** con un test positivo sono in realtà **sane** (falsi positivi).

Visualizzazione del Paradosso



🔍 Spiegazione del Paradosso

Il paradosso emerge perché la malattia è **così rara** che i **falsi positivi** tra le persone sane superano di gran lunga i **veri positivi** tra i malati:

≈1

Veri positivi

Su 10.000 persone, circa 1 è malata

99% di 1 persona = **≈1 vero positivo**

≈200

Falsi positivi

Su 10.000 persone, 9.999 sono sane

2% di 9.999 persone = **≈200 falsi positivi**

> 💡 **Lezione chiave:** La probabilità a priori **P(M) = 0.0001** (la bassa prevalenza della malattia) domina il risultato finale! Anche un test molto accurato produce molti falsi positivi quando applicato a una popolazione dove la condizione è rara.

Questo è il motivo per cui i medici raramente si basano su un singolo test per diagnosticare malattie rare, ma richiedono test di conferma o considerano altri fattori clinici.

Applicazioni a Machine Learning e AI

Il paradigma bayesiano - **Prior + Likelihood → Posterior** - è alla base di molti algoritmi fondamentali di intelligenza artificiale e machine learning. La capacità di aggiornare credenze con nuove evidenze è essenziale per sistemi intelligenti.

Naive Bayes Classifier

Obiettivo: Calcola **P(Classe | Caratteristiche)**

Formula: $P(C|x_1, ..., x_n) \propto P(C) \cdot P(x_1, ..., x_n|C)$

Applicazioni pratiche:

- Filtri spam per email
- Classificazione automatica di testi e documenti
- Analisi del sentiment su social media
- Categorizzazione di notizie

Il nome "Naive" deriva dall'assunzione di indipendenza condizionale tra le caratteristiche.

Inferenza Bayesiana

Obiettivo: Stima parametri da dati osservati

Formula: $P(\theta|\text{dati}) \propto P(\text{dati}|\theta) \cdot P(\theta)$

Esempi di utilizzo:

- Stima della probabilità di una moneta truccata
- Regressione bayesiana con incertezza quantificata
- Reti bayesiane per modellare dipendenze complesse
- A/B testing e ottimizzazione

A differenza dei metodi frequentisti, l'approccio bayesiano fornisce distribuzioni complete, non solo stime puntuali.

Aggiornamento Sequenziale

Paradigma: Aggiornare probabilità con nuove evidenze nel tempo

Principio chiave: Il posterior di oggi diventa il prior di domani

Applicazioni:

- Sistemi di raccomandazione adattivi
- Trading algoritmico e previsioni finanziarie
- Robotica e localizzazione (filtri di Kalman)
- Apprendimento online e continuo

Questo approccio è particolarmente potente quando i dati arrivano in modo incrementale o quando il contesto cambia nel tempo.

Il framework bayesiano offre un approccio unificato e matematicamente rigoroso per l'apprendimento automatico, combinando conoscenza a priori con evidenze empiriche.

Esempio Pratico: Filtro Spam Email

Vediamo come il teorema di Bayes viene applicato concretamente in uno dei sistemi più comuni di intelligenza artificiale: il filtraggio automatico delle email spam.

📧 Filtro Spam con Naive Bayes

Obiettivo del Sistema

Calcolare **P(Spam | Parole)** per ogni email in arrivo e classificarla automaticamente come spam o legittima.

Dati di Training

Il sistema ha analizzato migliaia di email già etichettate e ha calcolato:

- **P(Spam) = 0.3**
  Il 30% delle email sono spam (prior)
- **P("gratis" | Spam) = 0.8**
  L'80% delle email spam contengono "gratis"
- **P("gratis" | Non-Spam) = 0.1**
  Solo il 10% delle email legittime contengono "gratis"

Calcolo con Bayes

Riceviamo un'email che contiene la parola "gratis". Calcoliamo:

$$P(\text{Spam}|\text{"gratis"}) = \frac{P(\text{"gratis"}|\text{Spam}) \cdot P(\text{Spam})}{P(\text{"gratis"})}$$

Prima calcoliamo il denominatore con la probabilità totale:

P("gratis") = P("gratis"|Spam) · P(Spam) + P("gratis"|Non-Spam) · P(Non-Spam)

= (0.8 × 0.3) + (0.1 × 0.7) = 0.24 + 0.07 = 0.31

Quindi:

P(Spam | "gratis") = (0.8 × 0.3) / 0.31 = 0.24 / 0.31 ≈ **0.774**

77.4% probabilità di spam

Estensione a Multiple Parole

Nella pratica reale, i filtri spam considerano **multiple parole contemporaneamente**. Il classificatore Naive Bayes assume **indipendenza condizionale** tra le parole dato lo stato (spam/non-spam):

$$P(\text{Spam}|w_1, w_2, ..., w_n) \propto P(\text{Spam}) \cdot \prod_{i=1}^{n} P(w_i|\text{Spam})$$

Da qui deriva il nome **"Naive"** (ingenuo): l'assunzione di indipendenza è una semplificazione che raramente è vera nella realtà, ma funziona sorprendentemente bene in pratica!

Parole come "gratis", "vinci", "congratulazioni", "clicca qui" aumentano fortemente la probabilità di spam, mentre parole del contesto lavorativo la diminuiscono.

Errori Comuni e Misconcezioni

Quando si lavora con probabilità condizionali e il teorema di Bayes, ci sono tre errori fondamentali che vengono commessi frequentemente, anche da professionisti esperti. Comprenderli è essenziale per applicare correttamente questi concetti.

1

❌ Confondere P(A|B) con P(B|A)

L'errore: Pensare che P(A|B) = P(B|A) o che siano valori simili

Questa è probabilmente la confusione più comune e pericolosa. Le due probabilità condizionali sono generalmente **molto diverse**!

📊 Esempio: Pioggia e Nuvole

- **P(Nuvole | Pioggia) ≈ 1 (quasi 100%)**
  Quando piove, ci sono quasi sempre nuvole nel cielo
- **P(Pioggia | Nuvole) ≈ 0.3 (circa 30%)**
  Quando ci sono nuvole, piove solo circa il 30% delle volte

→ Queste sono probabilità **radicalmente diverse**! Una è quasi certa, l'altra è relativamente bassa.

2

❌ Ignorare la Probabilità a Priori

L'errore: Concentrarsi esclusivamente sulla verosimiglianza P(A|B) ignorando P(B)

Come abbiamo visto nell'esempio del test diagnostico, la probabilità a priori può avere un impatto **dominante** sul risultato finale.

📊 Esempio: Test Diagnostico

Nel test per la malattia rara che abbiamo analizzato:

- La **bassa prevalenza P(M) = 0.0001** (1 su 10.000) domina il calcolo
- Anche con un test accurato al 98-99%, P(M|Positivo) ≈ 0.49%
- Il prior basso produce molti falsi positivi assoluti, nonostante l'alta accuratezza relativa

→ La probabilità a priori è **cruciale** per interpretare correttamente qualsiasi evidenza!

Errori Comuni e Misconcezioni - Parte 2

1

❌ Prosecutor's Fallacy (Fallacia del Pubblico Ministero)

L'errore: Confondere P(Evidenza | Innocente) con P(Innocente | Evidenza)

Questo errore logico è particolarmente grave in contesti legali e forensi, dove può portare a condanne errate.

⚖️ Esempio Classico: Analisi DNA

Ragionamento SBAGLIATO (❌):

> "La probabilità che il DNA dell'imputato corrisponda per puro caso è 1 su 1 milione. Quindi la probabilità che l'accusato sia innocente è 1 su 1 milione."

Questo è un **errore logico grave**! Stiamo confondendo:

- **P(Match | Innocente) = 1/1.000.000**
  Probabilità di una corrispondenza casuale
- **P(Innocente | Match) = ?**
  Probabilità di innocenza data la corrispondenza

Ragionamento CORRETTO (✓):

Bisogna applicare il teorema di Bayes considerando anche:

- **P(Innocente)** a priori (quante persone potrebbero essere colpevoli?)
- **P(Match | Colpevole)** (probabilità di match se effettivamente colpevole)
- Altre evidenze e il contesto del caso

Se ci sono 10 milioni di persone nella popolazione di riferimento, ci aspetteremmo circa **10 matches casuali**, non solo uno!

⚠️ Regola d'Oro per Evitare Questi Errori

Usare sempre il teorema di Bayes esplicitamente quando si inverte una condizionalità, e non fare mai assunzioni intuitive!

Scrivere esplicitamente: $P(B|A) = [P(A|B) \cdot P(B)] / P(A)$

Calcolare ogni termine separatamente e verificare che abbia senso nel contesto.

Questi errori sono così comuni perché vanno contro la nostra intuizione naturale. La matematica delle probabilità condizionali richiede disciplina e rigore per essere applicata correttamente.

Riepilogo e Prossimi Passi

📋 Formule Fondamentali

Probabilità Condizionale

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

Base di tutto: come aggiornare credenze con nuove informazioni

Regola del Prodotto

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

Utile per calcolare intersezioni da probabilità condizionali

Probabilità Totale

$$P(A) = \sum_i P(A|B_i) \cdot P(B_i)$$

Scomporre probabilità complesse in scenari mutuamente esclusivi

Teorema di Bayes

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

Il cuore dell'inferenza bayesiana: invertire la condizionalità

🎯 Quando Usare Cosa?

01

P(A|B)

Quando vuoi aggiornare la probabilità di A avendo osservato B

02

Regola del Prodotto

Quando è più facile calcolare P(A|B) e P(B) separatamente che P(A ∩ B) direttamente

03

Probabilità Totale

Quando vuoi marginalizzare su una partizione dello spazio campionario

04

Teorema di Bayes

Quando devi invertire la condizionalità: passare da P(A|B) a P(B|A)

🚀 Applicazioni Pratiche Chiave

- **Machine Learning:** Naive Bayes, reti bayesiane, inferenza bayesiana
- **Medicina:** Interpretazione di test diagnostici e screening
- **Finanza:** Valutazione del rischio e decisioni di investimento
- **Data Science:** A/B testing, analisi predittiva, sistemi di raccomandazione

⚠️ Tre Errori Critici da Evitare

1. **Non confondere** P(A|B) con P(B|A) - sono generalmente molto diverse!
2. **Non ignorare** la probabilità a priori P(B) - può dominare il risultato
3. **Attenzione** al paradosso dei falsi positivi in test su popolazioni con bassa prevalenza

---

Prossima lezione: Approfondiremo le variabili aleatorie e le distribuzioni di probabilità, costruendo sulle fondamenta del ragionamento bayesiano che abbiamo stabilito oggi.