Lezione 3

Indipendenza Statistica e Condizionale

Quando gli eventi non si influenzano

Data Analyst e AI per Programmatori

Perché l'Indipendenza è Importante?

Semplificazione dei Calcoli

Quando due eventi A e B sono indipendenti, possiamo calcolare la probabilità della loro intersezione semplicemente moltiplicando le probabilità individuali:

> **$P(A \cap B) = P(A) \cdot P(B)$**

Fondamentale per il Machine Learning

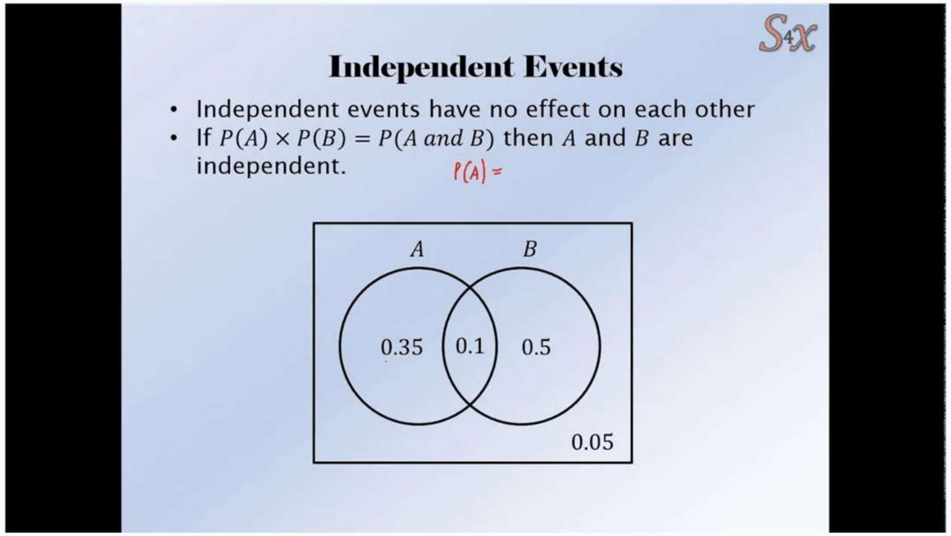
Molti algoritmi di ML si basano su assunzioni di indipendenza:

Naive Bayes: assume indipendenza condizionale delle feature

HMM: proprietà di Markov (indipendenza condizionale)

Reti Bayesiane:
sfruttano indipendenze per rappresentazione compatta

💡 **L'indipendenza riduce la complessità da esponenziale a lineare!**



Eventi indipendenti: sapere che B è accaduto non cambia P(A)

Definizione di Indipendenza Statistica

Definizione

Due eventi **A** e **B** sono **statisticamente indipendenti** se e solo se:

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

Interpretazione

Il verificarsi di **B** non influenza la probabilità di **A**, e viceversa. Sapere che uno degli eventi è accaduto non cambia la probabilità dell'altro.

Esempio: Eventi Indipendenti

Lancio di due monete:

• A = "prima moneta è testa"

• B = "seconda moneta è testa"

✓ Il risultato del primo lancio non influenza il secondo

Controesempio: Eventi Dipendenti

Estrazione carte senza reinserimento:

• A = "prima carta è un Re"

• B = "seconda carta è un Re"

✗ Se la prima è un Re, cambiano le probabilità per la seconda

Formulazioni Equivalenti dell'Indipendenza

Se $P(A) > 0$ e $P(B) > 0$, le seguenti affermazioni sono **equivalenti**:

1. Definizione Base

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

2. Probabilità Condizionale

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

3. Simmetrica

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

4. Regola del Prodotto

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

✓ Per verificare l'indipendenza, basta controllare UNA di queste condizioni

Tutte le formulazioni sono matematicamente equivalenti e portano alla stessa conclusione. In pratica, la formulazione (1) è spesso la più semplice da verificare quando si conoscono le probabilità individuali, mentre la formulazione (2) è utile quando si ragiona in termini di informazione.

Esempio: Lanci di Moneta Indipendenti

Setup del Problema

Lanciamo una moneta equa **due volte**. Definiamo:

A = "primo lancio è testa"

B = "secondo lancio è testa"

Spazio Campionario

Lancio 1	Lancio 2	Probabilità
T	T	**1/4**
T	C	1/4
C	T	1/4
C	C	1/4

A ∩ B = {(T, T)} evidenziato in giallo

Calcolo delle Probabilità

P(A) = P(primo lancio è T)

P(A) = 1/2

P(B) = P(secondo lancio è T)

P(B) = 1/2

P(A ∩ B) = P(entrambi T)

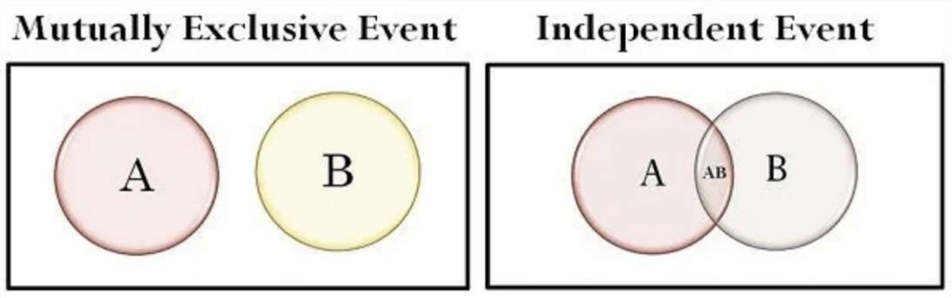
P(A ∩ B) = 1/4

✓ Verifica dell'Indipendenza

Controlliamo se P(A ∩ B) = P(A) · P(B):

1/4 = (1/2) · (1/2) = 1/4 ✓

⚠️ Attenzione: Disgiunti ≠ Indipendenti



Eventi Disgiunti

Due eventi A e B sono **disgiunti** se:

$$A \cap B = \emptyset \rightarrow P(A \cap B) = 0$$

Non possono accadere contemporaneamente.

Esempio: Lancio di un Dado

Sia A = "esce 1" e B = "esce 2"

- $P(A) = 1/6$
- $P(B) = 1/6$
- $P(A \cap B) = 0$ (disgiunti)
- **$P(A) \cdot P(B) = 1/36 \neq 0$**

Errore Comune

Molti studenti pensano che eventi disgiunti siano indipendenti. **Questo è FALSO!** Eventi disgiunti sono anzi **massimamente dipendenti**.

Conclusione

$P(A \cap B) = 0 \neq P(A) \cdot P(B) = 1/36$

→ A e B NON sono indipendenti!

Indipendenza a Coppie vs Indipendenza Mutua

Indipendenza a Coppie

Eventi $A_1, A_2, \ldots, A_n$ sono **indipendenti a coppie** se:

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$$

per ogni coppia $i \neq j$

Indipendenza Mutua (più forte)

Eventi $A_1, A_2, \ldots, A_n$ sono **mutuamente indipendenti** se per ogni sottoinsieme:

$$P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{ik}) = P(A_{i1}) \cdot P(A_{i2}) \cdot \ldots \cdot P(A_{ik})$$

⚠️ **Attenzione: Indipendenza a coppie NON implica indipendenza mutua!**

È possibile che eventi siano indipendenti a coppie ma non mutuamente indipendenti. Vediamo un esempio classico.

Esempio Classico (Bernstein): Due Monete Equa

Lanciamo due monete equa. Definiamo tre eventi:

Evento A

Prima moneta

Evento B

Seconda moneta

Evento C

Le due monete

Esempio di Bernstein: Verifica dell'Indipendenza

Spazio Campionario Completo:

Esito	A	B	C
TT	✓	✓	✓
TC	✓	✗	✗
CT	✗	✓	✗
CC	✗	✗	✓

Ricordiamo:

A: prima moneta è testa

B: seconda moneta è testa

C: le due monete mostrano lo stesso risultato

Verifica Indipendenza Mutua:

$P(A \cap B \cap C)$ $= P(TT) =$ **$1/4$**

$P(A) \cdot P(B) \cdot P(C)$ $= 1/2 \cdot 1/2 \cdot 1/2 =$ **$1/8$**

$1/4 \neq 1/8$ ✗

Conclusione

Gli eventi A, B, C sono **indipendenti a coppie** ma **NON mutuamente indipendenti**!

Spiegazione: *Se A e B accadono entrambi (entrambe teste), allora C accade automaticamente. Questo crea una dipendenza tra i tre eventi quando considerati insieme.*

Indipendenza Condizionale

Definizione

Due eventi **A** e **B** sono **condizionalmente indipendenti dato C** se:

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

Equivalentemente: $\mathbf{P(A \mid B \cap C) = P(A \mid C)}$

Interpretazione: Una volta che conosciamo C, sapere B non fornisce informazione aggiuntiva su A.

Notazione

$$A \perp B \mid C$$

"A è indipendente da B dato C"

⚠️ ATTENZIONE: Concetto Diverso dall'Indipendenza Statistica

Indipendenza Statistica

$A \perp B$ significa $P(A \cap B) = P(A) \cdot P(B)$

Indipendenza Condizionale

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

➤ A ⊥ B NON implica A ⊥ B | C (e viceversa!)

Indipendenza vs Indipendenza Condizionale

Concetto chiave: Indipendenza statistica (A ⊥ B) e indipendenza condizionale (A ⊥ B | C) sono concetti **diversi e indipendenti**. Possono verificarsi quattro situazioni:

A ⊥ B	A ⊥ B | C	Situazione	Esempio
✓	✓	**Caso 1:** Indipendenti sempre	Lanci di moneta indipendenti
✓	✗	**Caso 2:** Explaining away	Allarme-Terremoto-Ladro (collider)
✗	✓	**Caso 3:** Causa comune	Pioggia-Strada bagnata-Ombrelli (fork)
✗	✗	**Caso 4:** Dipendenza generale	Relazioni causali complesse

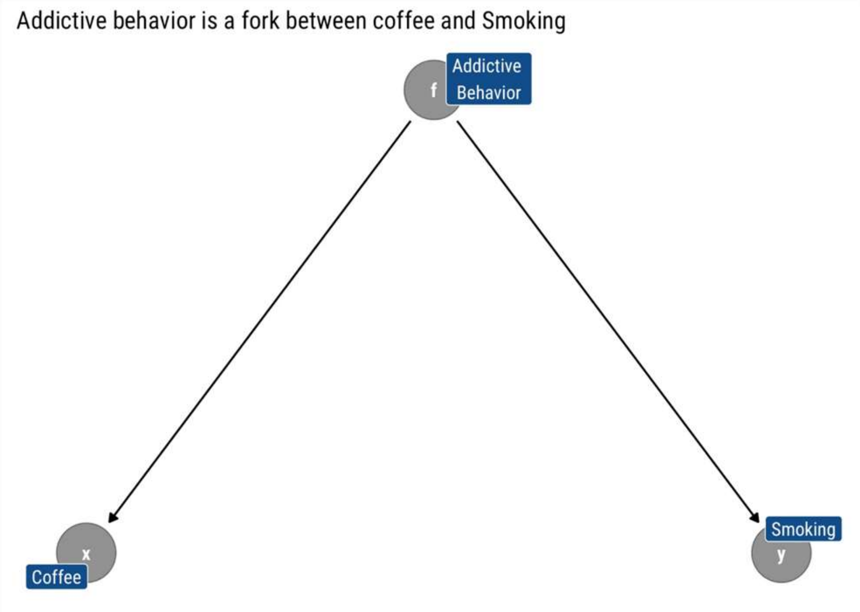
Caso 3: Causa Comune (più frequente)

Eventi dipendenti diventano indipendenti quando condizionati sulla loro causa comune. La causa "spiega" la dipendenza.

Caso 2: Explaining Away (controintuitivo)

Eventi indipendenti diventano dipendenti quando condizionati su un effetto comune. Una spiegazione "spiega via" l'altra.

Esempio: Causa Comune (Fork)



Struttura Fork: A ← C → B (C è causa comune)

Proprietà del Fork

Nella struttura fork, la causa comune C "spiega" la dipendenza tra A e B. Condizionando su C, A e B diventano indipendenti.

Scenario: La Pioggia

C: Piove

A: La strada è bagnata

B: Le persone hanno l'ombrello

Senza Condizionare su C:

A e B sono **dipendenti**: $A \not\perp B$

Vedere persone con ombrello aumenta la probabilità che la strada sia bagnata (entrambi causati dalla pioggia).

Condizionando su C (sappiamo che piove):

A e B diventano **indipendenti**: $A \perp B \mid C$

Se sappiamo che piove, sapere degli ombrelli non aggiunge informazione sulla strada. La causa comune spiega tutto.

Explaining Away (Collider)

Fenomeno Controintuitivo

Due eventi **indipendenti** diventano **dipendenti** quando condizionati su un effetto comune. Una spiegazione "spiega via" l'altra.

Struttura Causale: Collider

$$\text{Terremoto} \rightarrow \text{Allarme} \leftarrow \text{Ladro}$$

L'allarme è un **collider**: può essere causato da terremoto O da ladro. Questi sono eventi indipendenti.

Senza Condizionare

Terremoti e ladri sono eventi non correlati:

$$\text{Terremoto} \perp \text{Ladro}$$

Explaining Away: Esempio Numerico

Parametri del Modello

Probabilità a priori:

P(Terremoto) = 0.001

P(Ladro) = 0.01

Probabilità allarme:

P(A | T, L) = 1.0

P(A | T, ¬L) = 0.99

P(A | ¬T, L) = 0.95

P(A | ¬T, ¬L) = 0.001

Conclusione

Sapere che c'è un ladro "spiega via" l'allarme, rendendo il terremoto molto meno probabile.

Terremoto $\not\perp$ Ladro | Allarme

Risultati

Caso 1: Se l'allarme suona

P(Terremoto | Allarme) ≈ 9.4%

Caso 2: Se allarme E ladro

P(Terremoto | Allarme, Ladro) ≈ 0.1%

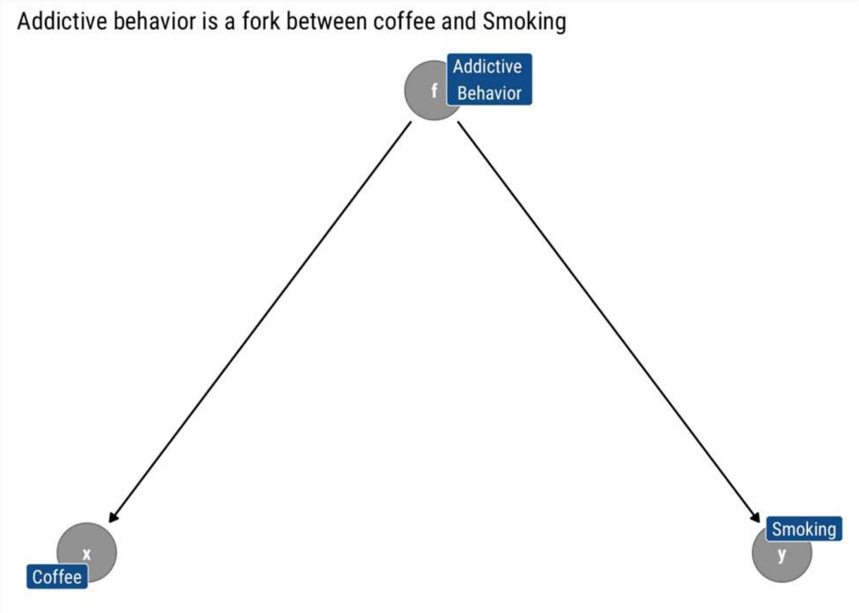
Interpretazione

La probabilità di terremoto diminuisce drasticamente (da 9.4% a 0.1%) quando scopriamo che c'è un ladro. Il ladro "spiega" l'allarme, rendendo il terremoto improbabile.

Applicazioni: Reti Bayesiane

Cosa sono le Reti Bayesiane?

Le **reti bayesiane** sono grafi diretti aciclici che rappresentano dipendenze probabilistiche tra variabili. Sfruttano l'**indipendenza condizionale** per rappresentare distribuzioni complesse in modo compatto ed efficiente.



Tre Strutture Fondamentali

1. Chain (Catena): $A \rightarrow B \rightarrow C$

$A \perp C \mid B$ (B "scherma" A da C)

2. Fork (Causa Comune): $A \leftarrow B \rightarrow C$

$A \perp C \mid B$ (B è la causa comune)

3. Collider (Effetto Comune): $A \rightarrow B \leftarrow C$

$A \perp C$ (non condizionato), $A \not\perp C \mid B$

Riduzione della Complessità

Senza indipendenze, rappresentare la distribuzione congiunta di **n** variabili binarie richiede:

Esempio Pratico

Per 10 variabili binarie:

Senza indipendenze: $2^{10} - 1 = 1023$ parametri

Applicazioni al Machine Learning

1. Naive Bayes Classifier

$P(X_1, X_2, ..., X_n \mid Y) =$

$\prod_i P(X_i \mid Y)$

Assunzione: Le feature sono condizionalmente indipendenti data la classe.

Applicazioni: Classificazione testi, spam filter, sentiment analysis

2. Hidden Markov Models

$P(S_{t+1} \mid S_1, ..., S_t) =$

$P(S_{t+1} \mid S_t)$

Proprietà di Markov: Il futuro dipende solo dal presente, non dal passato.

Applicazioni: Speech recognition, POS tagging, time series

3. Graphical Models

Rappresentazione compatta

$P(X_1, ..., X_n) = \prod_i P(X_i \mid Pa(X_i))$

Struttura: Sfruttano indipendenze condizionali per fattorizzare la distribuzione.

Tipi: Bayesian Networks, Markov Random Fields, CRF

Benefici dell'Indipendenza Condizionale

Riduzione Parametri

Da $O(2^n)$ a $O(n)$ parametri per n variabili binarie

Inferenza Efficiente

Calcoli trattabili anche con molte variabili

Interpretabilità

Struttura causale esplicita e comprensibile

💡 **Perché funziona?**

Anche quando l'assunzione di indipendenza condizionale è violata, questi modelli spesso

⚙️ **Trade-off**

Sacrifichiamo accuratezza delle probabilità assolute per guadagnare efficienza

Riepilogo e Prossimi Passi

Concetti Chiave della Lezione 3

Indipendenza Statistica

- $P(A \cap B) = P(A) \cdot P(B)$
- Equivalente a $P(A \mid B) = P(A)$
- Eventi disgiunti NON sono indipendenti
- Indipendenza a coppie ≠ indipendenza mutua

Indipendenza Condizionale

- $P(A \cap B \mid C) = P(A \mid C) \cdot P(B \mid C)$
- Notazione: $A \perp B \mid C$
- Concetto diverso da $A \perp B$
- $A \perp B$ NON implica $A \perp B \mid C$

Strutture Causali

- **Chain:** $A \to B \to C$ ($A \perp C \mid B$)
- **Fork:** $A \leftarrow B \to C$ ($A \perp C \mid B$)
- **Collider:** $A \to B \leftarrow C$ ($A \perp C$, ma $A \not\perp C \mid B$)

Applicazioni Pratiche

- Naive Bayes: assume $A \perp B \mid Y$
- Reti Bayesiane: sfruttano indipendenze
- HMM: proprietà di Markov
- Semplificazione dei calcoli

Prossima Lezione: Distribuzioni di Probabilità Discrete

- Variabili aleatorie discrete
- Funzioni di massa
- Bernoulli e Binomiale
- Geometrica
- Modellazione eventi
- Analisi dati categorici