

Probabilità e inferenza: dalla formula di Bayes alle reti neurali

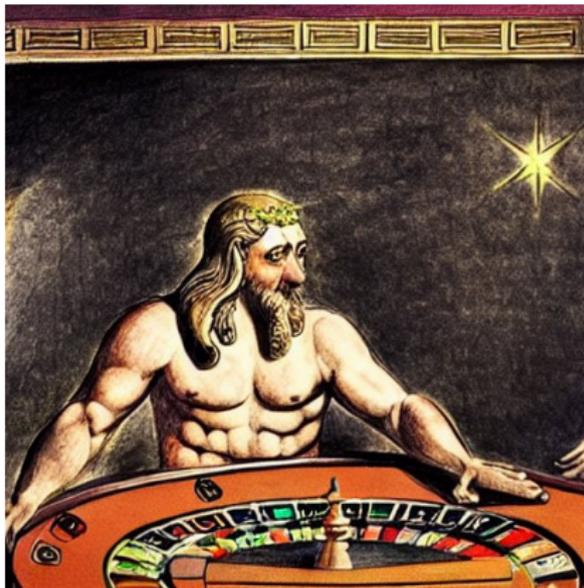
Dario Trevisan

Università di Pisa
dario.trevisan@unipi.it

Introduzione

- Il **calcolo probabilità** è spesso associato al gioco d'azzardo e alle scommesse
- Storicamente è nata proprio per queste soddisfare le **esigenze** degli scommettitori!
- Se avete già incontrato la probabilità nei vostri studi, vi sarà stato chiesto di rispondere a domande su *lanci di dadi, monete, estrazione di carte, ecc.*





In realtà, le applicazioni della probabilità sono ben più diffuse e **importanti**:

- nelle previsioni meteorologiche e **climatiche**
- in **medicina**
- nell'economia e la finanza
- in tutte le scienze sperimentali (dalla fisica atomica, alla psicologia)



- E la **matematica**?
- È l'apice del **ragionamento logico**, non c'è spazio (in apparenza) per il caso!
- La probabilità estende rigore matematico alle situazioni reali (in cui c'è incertezza).

- Facendo quindi un'**analogia**:
- Il **gioco d'azzardo** è per la probabilità quello che un **laboratorio** è per una scienza sperimentale.
- **NON** tutti i matematici che lavorano nella probabilità sono scommettitori!



Il ragionamento deduttivo e i suoi limiti

La **logica** codifica delle regole per **dedurre** la validità di affermazioni a partire da altre (già valide).

Le affermazioni si indicano astrattamente con lettere A, B, C, \dots

Esempi:

$A =$ “Oggi piove”

$B =$ “Mario porta l’ombrello”

$C =$ “ $2 + 2 = 5$ ”



Una regola molto comune è il **modus ponens** (sillogismo).

- Date affermazioni A , B ,
- se valgono le due affermazioni

$$\left\{ \begin{array}{l} \text{“se } A, \text{ allora } B\text{”} \\ A, \end{array} \right.$$

- si può dedurre che vale anche B .

L'affermazione “se A , allora B ”, si interpreta

“ogni volta che A è vera, è vera anche B ”.

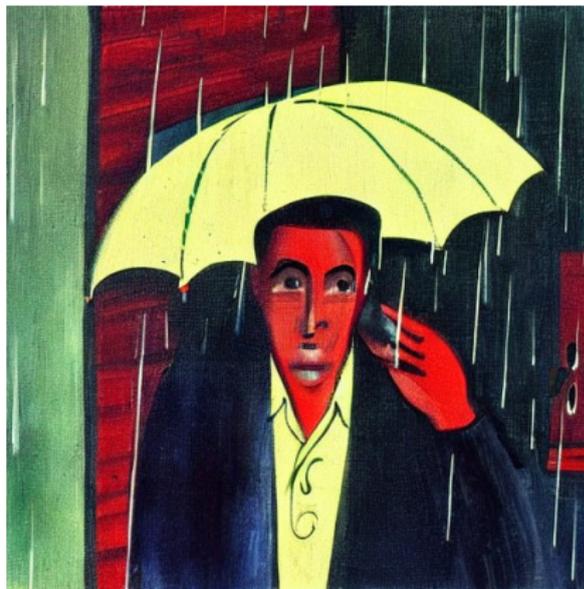
Esempio:

- conosciamo bene Mario, che è una persona molto attenta, e ogni volta che piove, esce con l'ombrello ("se A allora B")
- osserviamo che oggi piove (A)
- deduciamo che Mario è sicuramente uscito con l'ombrello (anche senza averlo incontrato)



Esempio:

- conosciamo bene Mario, che è una persona molto attenta, e **ogni volta che piove, esce con l'ombrello** (“se A allora B ”)
- osserviamo che **oggi piove** (A)
- **deduciamo** che Mario è sicuramente uscito con l'ombrello (anche senza averlo incontrato)



Tutto il **sapere matematico** è costruito (con fatica!) partendo da poche premesse valide (gli **assiomi**) e usando regole come queste per dedurre altre verità (i **teoremi**).

Ma la logica non contempla tutti i modi **ragionevoli** di pensare, in particolare:

- Date affermazioni A , B ,
- se valgono le due affermazioni

$$\left\{ \begin{array}{l} \text{“se } A, \text{ allora } B\text{”} \\ B, \end{array} \right.$$

- cosa si può dedurre su A ?

Nell'esempio: supponiamo di vedere Mario che cammina con l'ombrello, possiamo dedurre che **necessariamente** piove?

- Ovviamente non possiamo **dedurre** che piove.
- E se invece di vedere solo Mario con l'ombrello, tutte le persone per strada ce l'hanno?
- Ancora non possiamo dedurre che piove, ma dovrebbe essere **più plausibile**.



Valori di verità e probabilità

- Un'affermazione A è caratterizzata dal poter attribuire ad essa (almeno in linea di principio) un **valore di verità**

Vero (1), Falso (0).

- Tale valore può dipendere dal contesto, in particolare tutta l'informazione di cui si dispone e che si ritiene vera (ad esempio, gli assiomi e tutti i teoremi già dimostrati): indichiamola con la lettera I .

Valori di verità e probabilità

- Un'affermazione A è caratterizzata dal poter attribuire ad essa (almeno in linea di principio) un **valore di verità**

Vero (1), Falso (0).

- Tale valore può dipendere dal contesto, in particolare tutta l'informazione di cui si dispone e che si ritiene vera (ad esempio, gli assiomi e tutti i teoremi già dimostrati): indichiamola con la lettera I .
- Anche se l'informazione I non basta a **dedurre** il valore di verità di A , attribuiamo un **grado di fiducia**

$$P(A|I)$$

che si chiama la **probabilità** di A sapendo I (o condizionata ad I).

- Richiediamo che

$$0 \leq P(A|I) \leq 1.$$

- Gli estremi recuperano i valori di verità:

$P(A|I) = 1$ significa che A è ritenuta **vera**, sapendo I ,

$P(A|I) = 0$ significa che A è ritenuta **falsa**, sapendo I .

- Richiediamo che

$$0 \leq P(A|I) \leq 1.$$

- Gli estremi recuperano i valori di verità:

$P(A|I) = 1$ significa che A è ritenuta **vera**, sapendo I ,

$P(A|I) = 0$ significa che A è ritenuta **falsa**, sapendo I .

- Nella pratica conviene evitare di attribuire i valori 0 o 1, altrimenti **nessuna evidenza**, per quanto importante, potrà mai **cambiare** tali valori (seguendo le regole di calcolo).
- È anche difficile distinguere **intuitivamente** la differenza tra probabilità piccole e molto piccole, ad esempio $10^{-6} > 10^{-9}$, ma entrambe ci sembrano piccole.

Regola della somma

- Dalla logica: se A è vera, la sua negazione, “non A ” è falsa.
- Quindi esattamente una tra A e “non A ” deve essere vera,
- I due valori di verità quindi sommano sempre a 1.

Regola della somma

- Dalla logica: se A è vera, la sua **negazione**, “non A ” è falsa.
- Quindi esattamente una tra A e “non A ” deve essere vera,
- I due **valori di verità** quindi sommano sempre a 1.
- Questo motiva l'estensione alle probabilità tramite la **regola della somma**

$$P(A|I) + P(\text{non } A|I) = 1.$$



Nelle scommesse è comune usare le quote (in inglese, **odds**):

$$\text{Odds}(A|I) = P(A|I) : P(\text{non } A|I),$$

che indicano l'inverso di quanto denaro si **guadagna** in caso vincita (scommettendo 1€).

Nelle scommesse è comune usare le quote (in inglese, **odds**):

$$\text{Odds}(A|I) = P(A|I) : P(\text{non } A|I),$$

che indicano l'inverso di quanto denaro si **guadagna** in caso vincita (scommettendo 1€). **Esempi:**

- $\text{Odds}(A|I) = 1 : 1$ significa che scommettendo 1€, in caso di vincita si guadagna 1€ (quindi ci si ritrova con 2€ in tasca). La probabilità è $P(A|I) = 1/2 = 50\%$.

Nelle scommesse è comune usare le quote (in inglese, **odds**):

$$\text{Odds}(A|I) = P(A|I) : P(\text{non } A|I),$$

che indicano l'inverso di quanto denaro si **guadagna** in caso vincita (scommettendo 1€). **Esempi:**

- $\text{Odds}(A|I) = 1 : 1$ significa che scommettendo 1€, in caso di vincita si guadagna 1€ (quindi ci si ritrova con 2€ in tasca). La probabilità è $P(A|I) = 1/2 = 50\%$.
- $\text{Odds}(A|I) = 1 : 2$ significa che scommettendo 1€, in caso di vincita si guadagna 2€ (quindi ci si ritrova con 3€ in tasca). La probabilità è $P(A|I) = 1/3 \approx 33\%$.

Nelle scommesse è comune usare le quote (in inglese, **odds**):

$$\text{Odds}(A|I) = P(A|I) : P(\text{non } A|I),$$

che indicano l'inverso di quanto denaro si **guadagna** in caso vincita (scommettendo 1€). **Esempi:**

- $\text{Odds}(A|I) = 1 : 1$ significa che scommettendo 1€, in caso di vincita si guadagna 1€ (quindi ci si ritrova con 2€ in tasca). La probabilità è $P(A|I) = 1/2 = 50\%$.
- $\text{Odds}(A|I) = 1 : 2$ significa che scommettendo 1€, in caso di vincita si guadagna 2€ (quindi ci si ritrova con 3€ in tasca). La probabilità è $P(A|I) = 1/3 \approx 33\%$.
- $\text{Odds}(A|I) = 3 : 2 = 1 : (2/3)$ significa che scommettendo 1€, in caso di vincita si guadagna $2/3$ € (quindi ci si ritrova con $5/3$ € in tasca), oppure che scommettendo 3 €, in caso di vincita si guadagna 2€.

Regola del prodotto

- In logica, la **congiunzione** “ A e B ” tra due affermazioni è vera solo quando **entrambe** sono vere.
- Il valore di verità di “ A e B ” è il prodotto dei due valori di verità.

Regola del prodotto

- In logica, la **congiunzione** “ A e B ” tra due affermazioni è vera solo quando **entrambe** sono vere.
- Il valore di verità di “ A e B ” è il prodotto dei due valori di verità.
- Possiamo anche procedere così:
 - **prima** si calcola il valore di verità di A
 - **poi**, se A è vera, si calcola il valore di verità di B .

Regola del prodotto

- In logica, la **congiunzione** “A e B” tra due affermazioni è vera solo quando **entrambe** sono vere.
- Il valore di verità di “A e B” è il prodotto dei due valori di verità.
- Possiamo anche procedere così:
 - **prima** si calcola il valore di verità di A
 - **poi**, se A è vera, si calcola il valore di verità di B.
- Questo motiva in probabilità l'introduzione della **regola del prodotto**:

$$P(A \text{ e } B|I) = P(A|I) \cdot P(B|A \text{ e } I).$$

Esempio:

- La probabilità che oggi piova (A) è $P(A|I) = 10\%$.
- La probabilità che Mario porti l'ombrello (B), sapendo che fuori piove, è

$$P(B|A \text{ e } I) = 95\%.$$

- La probabilità che oggi piova e Mario porti l'ombrello (A e B) è

$$P(A \text{ e } B|I) = 10\% \cdot 95\% = 0,095.$$

Una prima conseguenza è la **monotonia** della probabilità:

$$P(A \text{ e } B|I) = P(A|I) \cdot P(B|A \text{ e } I) \leq P(A|I).$$

Una prima conseguenza è la **monotonia** della probabilità:

$$P(A \text{ e } B|I) = P(A|I) \cdot P(B|A \text{ e } I) \leq P(A|I).$$

Esempio:

- Linda ha 31 anni, è single, schietta e molto brillante.
- Ha conseguito la laurea in filosofia.
- Da studentessa, era preoccupata per i problemi di discriminazione e giustizia sociale.
- Ha anche partecipato a dimostrazioni anti-nucleari.



Una prima conseguenza è la **monotonia** della probabilità:

$$P(A \text{ e } B|I) = P(A|I) \cdot P(B|A \text{ e } I) \leq P(A|I).$$



Esempio:

- Linda ha 31 anni, è single, schietta e molto brillante.
- Ha conseguito la laurea in filosofia.
- Da studentessa, era preoccupata per i problemi di discriminazione e giustizia sociale.
- Ha anche partecipato a dimostrazioni anti-nucleari.

Quale è più probabile?

- (1) Linda è una cassiera di banca.
- (2) Linda è una cassiera di banca ed è attiva nel movimento femminista.

La regola del prodotto recupera il **modus ponens** della logica.

- Date affermazioni A , B , informazione nota I ,
- A è vera si traduce in probabilità con $P(A|I) = 1$,
- “se A allora B ” è vera si traduce con $P(B|A \text{ e } I) = 1$,

La regola del prodotto recupera il **modus ponens** della logica.

- Date affermazioni A , B , informazione nota I ,
- A è vera si traduce in probabilità con $P(A|I) = 1$,
- “se A allora B ” è vera si traduce con $P(B|A \text{ e } I) = 1$,
- dalla regola del prodotto (e la monotonia), si trova

$$P(B|I) \geq P(A \text{ e } B|I) = P(A|I) \cdot P(B|A \text{ e } I) = 1 \cdot 1 = 1.$$

Formula di Bayes

Ma la conseguenza più importante è una versione della (non-)regola

$$\left\{ \begin{array}{l} \text{"se } A, \text{ allora } B" \\ B \end{array} \right. \Rightarrow A?$$

Formula di Bayes

Ma la conseguenza più importante è una versione della (non-)regola

$$\left\{ \begin{array}{l} \text{"se } A, \text{ allora } B" \\ B \end{array} \right. \Rightarrow A?$$

Date affermazioni A , B ,

Formula di Bayes

Ma la conseguenza più importante è una versione della (non-)regola

$$\left\{ \begin{array}{l} \text{"se } A, \text{ allora } B" \\ B \end{array} \right. \Rightarrow A?$$

Date affermazioni A , B ,

- Le affermazioni " A e B ", " B e A " sono logicamente equivalenti, quindi

$$P(A \text{ e } B|I) = P(B \text{ e } A|I)$$

Formula di Bayes

Ma la conseguenza più importante è una versione della (non-)regola

$$\left\{ \begin{array}{l} \text{"se } A, \text{ allora } B" \\ B \end{array} \right. \Rightarrow A?$$

Date affermazioni A , B ,

- Le affermazioni “ A e B ”, “ B e A ” sono logicamente equivalenti, quindi

$$P(A \text{ e } B|I) = P(B \text{ e } A|I)$$

- Applichiamo la regola del prodotto ad entrambi i membri:

$$P(A|I)P(B|A \text{ e } I) = P(B|I)P(A|B \text{ e } I)$$

Formula di Bayes

Ma la conseguenza più importante è una versione della (non-)regola

$$\left\{ \begin{array}{l} \text{"se } A, \text{ allora } B" \\ B \end{array} \right. \Rightarrow A?$$

Date affermazioni A , B ,

- Le affermazioni “ A e B ”, “ B e A ” sono logicamente equivalenti, quindi

$$P(A \text{ e } B|I) = P(B \text{ e } A|I)$$

- Applichiamo la regola del prodotto ad entrambi i membri:

$$P(A|I)P(B|A \text{ e } I) = P(B|I)P(A|B \text{ e } I)$$

- Isoliamo il termine più a destra dividendo per $P(B|I)$ (> 0):

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

Riscriviamola (sottointendendo I per chiarezza):

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

Riscriviamola (sottointendendo I per chiarezza):

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

- $P(A)$ è detta probabilità **a priori** (prima di sapere che B è vero)

Riscriviamola (sottointendendo I per chiarezza):

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

- $P(A)$ è detta probabilità **a priori** (prima di sapere che B è vero)
- $P(B|A)$ è detta **verosimiglianza** (Likelihood in inglese) di A relativa a B ,

$$L(A; B) = P(B|A)$$

Riscriviamola (sottointendendo I per chiarezza):

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

- $P(A)$ è detta probabilità **a priori** (prima di sapere che B è vero)
- $P(B|A)$ è detta **verosimiglianza** (Likelihood in inglese) di A relativa a B ,

$$L(A; B) = P(B|A)$$

- la probabilità $P(A|B)$ è detta **a posteriori** (sapendo che B è vero).

Riscriviamola (sottointendendo I per chiarezza):

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

- $P(A)$ è detta probabilità **a priori** (prima di sapere che B è vero)
- $P(B|A)$ è detta **verosimiglianza** (Likelihood in inglese) di A relativa a B ,

$$L(A; B) = P(B|A)$$

- la probabilità $P(A|B)$ è detta **a posteriori** (sapendo che B è vero).
- il rapporto

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

può essere

- maggiore di 1 (in quel caso $P(A|B) > P(A)$),
- minore di 1 (e allora $P(A|B) < P(A)$)
- uguale ad 1 (allora $P(A|B) = P(A)$).

Passando alle **odds**, l'interpretazione della formula di Bayes è più chiara:

- Scriviamo le formule

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

$$P(\text{non } A|B) = P(\text{non } A) \cdot \frac{L(\text{non } A; B)}{P(B)}$$

Passando alle **odds**, l'interpretazione della formula di Bayes è più chiara:

- Scriviamo le formule

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

$$P(\text{non } A|B) = P(\text{non } A) \cdot \frac{L(\text{non } A; B)}{P(B)}$$

- Se le dividiamo, $P(B)$ si semplifica e troviamo

$$\text{Odds}(A|B) = \text{Odds}(A) \cdot \frac{L(A; B)}{L(\text{non } A; B)}$$

Passando alle **odds**, l'interpretazione della formula di Bayes è più chiara:

- Scriviamo le formule

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

$$P(\text{non } A|B) = P(\text{non } A) \cdot \frac{L(\text{non } A; B)}{P(B)}$$

- Se le dividiamo, $P(B)$ si semplifica e troviamo

$$\text{Odds}(A|B) = \text{Odds}(A) \cdot \frac{L(A; B)}{L(\text{non } A; B)}$$

- Le odds di A a posteriori (ossia sapendo B vera) si ottengono moltiplicando le odds di A a priori per il **rapporto di verosimiglianza**

$$\frac{L(A; B)}{L(\text{non } A; B)} = \frac{P(B|A)}{P(B|\text{non } A)}$$

Esempio: $A = \text{“Piove”}$, $B = \text{“Mario prende l’ombrello”}$.

Esempio: $A = \text{"Piove"}$, $B = \text{"Mario prende l'ombrello"}$.

- $P(\text{Piove}) = 10\%$, quindi

$$\text{Odds}(\text{Piove}) = 1 : 9,$$

- $P(\text{Mario prende l'ombrello} | \text{Piove}) = 95\%$, quindi

$$L(\text{Piove}; \text{Mario prende l'ombrello}) = 95\%,$$

- $P(\text{Mario prende l'ombrello} | \text{non Piove}) = 2\%$, quindi

$$L(\text{non Piove}; \text{Mario prende l'ombrello}) = 2\%.$$

- Ne segue che

$$\text{Odds}(\text{Piove} | \text{Mario prende l'ombrello}) = \frac{1}{9} \cdot \frac{95\%}{2\%} \approx 5.3,$$

è molto probabile che stia piovendo, se incontriamo Mario con l'ombrello!

$$\text{Odds}(A|B) = \text{Odds}(A) \cdot \frac{L(A; B)}{L(\text{non } A; B)}.$$

$$\text{Odds}(A|B) = \text{Odds}(A) \cdot \frac{L(A; B)}{L(\text{non } A; B)}.$$

$L(A; B)$ **da sola** non è sufficiente, va confrontata con $L(\text{non } A; B)$.

- Se in una scena del crimine si trova un indizio (B) che è una prova
 - **tanto** a favore della colpevolezza di un sospettato (A)
 - **quanto** della sua innocenza ($\text{non } A$)

allora è inutile!

$$\text{Odds}(A|B) = \text{Odds}(A) \cdot \frac{L(A; B)}{L(\text{non } A; B)}.$$

$L(A; B)$ **da sola** non è sufficiente, va confrontata con $L(\text{non } A; B)$.

- Se in una scena del crimine si trova un indizio (B) che è una prova
 - **tanto** a favore della colpevolezza di un sospettato (A)
 - **quanto** della sua innocenza ($\text{non } A$)

allora è inutile!

Le odds **a priori** $\text{Odds}(A)$ sono pure importanti:

- se sono molto piccole (A è poco probabile), affinché $\text{Odds}(A|B)$ sia grande (ad esempio, ≥ 1) è necessario che il rapporto di verosimiglianza sia molto grande:

“Affermazioni straordinarie richiedono prove straordinarie”.

- Incontrate per strada una persona che vi dice

“nel garage ho un'automobile”

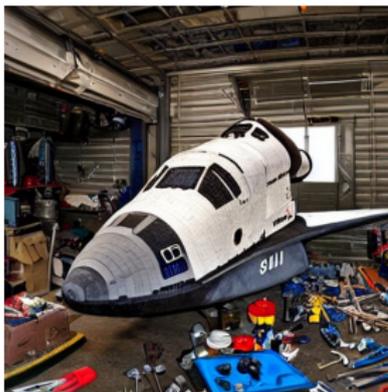
- e vi mostra ha un paio di chiavi.
- Sareste disposti a credergli?



- Incontrate per strada una persona che vi dice

“nel garage ho un'automobile”

- e vi mostra ha un paio di chiavi.
- Sareste disposti a credergli?



- E se invece vi dicesse
“nel garage ho un'astronave”
- basterebbe il paio di chiavi a convincervi?

Il tacchino induttivista



Un tacchino notò che fin dal primo giorno nell'allevamento il cibo arrivava puntuale alle 9 del mattino, di mercoledì e di giovedì, nei giorni caldi e nei giorni freddi, sia che piovesse sia che splendesse il sole.

*Allora ne inferì: "Mi danno **sempre** il cibo alle 9 del mattino".*

Il tacchino induttivista



Un tacchino notò che fin dal primo giorno nell'allevamento il cibo arrivava puntuale alle 9 del mattino, di mercoledì e di giovedì, nei giorni caldi e nei giorni freddi, sia che piovesse sia che splendesse il sole.

*Allora ne inferì: "Mi danno **sempre** il cibo alle 9 del mattino".*

*Questa concezione si rivelò decisamente falsa alla **vigilia di Natale**...*



- $A =$ “Mi danno **sempre** il cibo alle 9 del mattino” (il padrone non mi ucciderà mai).
- $B =$ “Oggi il padrone mi porta il cibo alle 9 del mattino” (“non B ” significa che il padrone viene invece ad ucciderlo).

- A = “Mi danno **sempre** il cibo alle 9 del mattino” (il padrone non mi ucciderà mai).
- B = “Oggi il padrone mi porta il cibo alle 9 del mattino” (“non B ” significa che il padrone viene invece ad ucciderlo).
- Poniamo

$$\begin{cases} L(A; B) & = P(B|A) = 1 \\ L(\text{non } A; B) & = P(B|\text{non } A) = 99\% \end{cases} \Rightarrow \frac{L(A; B)}{L(\text{non } A; B)} \approx 1.01.$$

- Dopo un giorno in cui il tacchino riceve il cibo le odds a posteriori sono

$$\text{Odds}(A|B) \approx \text{Odds}(A) \cdot (1.01).$$

- A = “Mi danno **sempre** il cibo alle 9 del mattino” (il padrone non mi ucciderà mai).
- B = “Oggi il padrone mi porta il cibo alle 9 del mattino” (“non B ” significa che il padrone viene invece ad ucciderlo).
- Poniamo

$$\begin{cases} L(A; B) & = P(B|A) = 1 \\ L(\text{non } A; B) & = P(B|\text{non } A) = 99\% \end{cases} \Rightarrow \frac{L(A; B)}{L(\text{non } A; B)} \approx 1.01.$$

- Dopo un giorno in cui il tacchino riceve il cibo le odds a posteriori sono

$$\text{Odds}(A|B) \approx \text{Odds}(A) \cdot (1.01).$$

- Ripetendo, dopo n giorni che riceve il cibo sono

$$\text{Odds}(A|B_n) \approx \text{Odds}(A) \cdot (1.01)^n.$$

- Per n grande, il termine $(1.01)^n$ diventa esponenzialmente **grande**.

- A = “Mi danno **sempre** il cibo alle 9 del mattino” (il padrone non mi ucciderà mai).
- B = “Oggi il padrone mi porta il cibo alle 9 del mattino” (“non B ” significa che il padrone viene invece ad ucciderlo).
- Poniamo

$$\begin{cases} L(A; B) & = P(B|A) = 1 \\ L(\text{non } A; B) & = P(B|\text{non } A) = 99\% \end{cases} \Rightarrow \frac{L(A; B)}{L(\text{non } A; B)} \approx 1.01.$$

- Dopo un giorno in cui il tacchino riceve il cibo le odds a posteriori sono

$$\text{Odds}(A|B) \approx \text{Odds}(A) \cdot (1.01).$$

- Ripetendo, dopo n giorni che riceve il cibo sono

$$\text{Odds}(A|B_n) \approx \text{Odds}(A) \cdot (1.01)^n.$$

- Per n grande, il termine $(1.01)^n$ diventa esponenzialmente **grande**.
- Però $\text{Odds}(A)$ dovrebbe essere **piccola**, se il tacchino sapesse ad esempio della sorte che è toccata a tutti i suoi colleghi. . .

Sistemi di alternative e decisioni

Sistemi di alternative e decisioni

- Un **sistema di alternative** è una famiglia di affermazioni

$$A_1, A_2, A_3, \dots, A_n,$$

tali che, assumendo l'informazione I , si sa che **una e una sola** è vera.

Sistemi di alternative e decisioni

- Un **sistema di alternative** è una famiglia di affermazioni

$$A_1, A_2, A_3, \dots, A_n,$$

tali che, assumendo l'informazione I , si sa che **una e una sola** è vera.

- **Esempio:** si lancia un dado a 6 facce:

$$A_1 = \text{"l'esito del lancio è 1"}, A_2 = \text{"l'esito del lancio è 2"}, \dots$$

Sistemi di alternative e decisioni

- Un **sistema di alternative** è una famiglia di affermazioni

$$A_1, A_2, A_3, \dots, A_n,$$

tali che, assumendo l'informazione I , si sa che **una e una sola** è vera.

- **Esempio:** si lancia un dado a 6 facce:

$$A_1 = \text{"l'esito del lancio è 1"}, A_2 = \text{"l'esito del lancio è 2"}, \dots$$

- Data A , qualsiasi, possiamo completare ad un sistema con "non A ".

Sistemi di alternative e decisioni

- Un **sistema di alternative** è una famiglia di affermazioni

$$A_1, A_2, A_3, \dots, A_n,$$

tali che, assumendo l'informazione I , si sa che **una e una sola** è vera.

- **Esempio:** si lancia un dado a 6 facce:

$$A_1 = \text{"l'esito del lancio è 1"}, A_2 = \text{"l'esito del lancio è 2"}, \dots$$

- Data A , qualsiasi, possiamo completare ad un sistema con "non A ".
- La regola della somma si estende ai sistemi di alternative:

$$P(A_1|I) + P(A_2|I) + \dots + P(A_n|I) = 1.$$

Sistemi di alternative e decisioni

- Un **sistema di alternative** è una famiglia di affermazioni

$$A_1, A_2, A_3, \dots, A_n,$$

tali che, assumendo l'informazione I , si sa che **una e una sola** è vera.

- Esempio:** si lancia un dado a 6 facce:

$$A_1 = \text{"l'esito del lancio è 1"}, A_2 = \text{"l'esito del lancio è 2"}, \dots$$

- Data A , qualsiasi, possiamo completare ad un sistema con "non A ".
- La regola della somma si estende ai sistemi di alternative:

$$P(A_1|I) + P(A_2|I) + \dots + P(A_n|I) = 1.$$

- Più in generale, preso un sottoinsieme $J \subseteq \{1, 2, \dots, n\}$,

$$P(\text{vale una tra le } A_j, j \in J|I) = \sum_{j \in J} P(A_j|I).$$

- Come attribuire le probabilità

$$P(A_1|I), \quad P(A_2|I), \quad \dots, \quad P(A_n|I)$$

di un sistema di alternative?

- Come attribuire le probabilità

$$P(A_1|I), \quad P(A_2|I), \quad \dots, \quad P(A_n|I)$$

di un sistema di alternative?

- Dipende dall'informazione I di cui si dispone (che si ritiene vera)

- Come attribuire le probabilità

$$P(A_1|I), \quad P(A_2|I), \quad \dots, \quad P(A_n|I)$$

di un sistema di alternative?

- Dipende dall'informazione I di cui si dispone (che si ritiene vera)
- Nel caso del lancio del dado, **prima** di osservare l'esito, siamo tentati di attribuire

$$P(A_1|I) = P(A_2|I) = \dots = P(A_6|I) = \frac{1}{6}.$$

- Come attribuire le probabilità

$$P(A_1|I), \quad P(A_2|I), \quad \dots, \quad P(A_n|I)$$

di un sistema di alternative?

- Dipende dall'informazione I di cui si dispone (che si ritiene vera)
- Nel caso del lancio del dado, **prima** di osservare l'esito, siamo tentati di attribuire

$$P(A_1|I) = P(A_2|I) = \dots = P(A_6|I) = \frac{1}{6}.$$

- Laplace formulò il **principio di indifferenza** (o di ragione insufficiente):

In assenza di evidenze che permettano di preferire un'alternativa ad un'altra, si deve assegnare la medesima probabilità a tutte le alternative.

- Se la probabilità di tutte le alternative è la stessa,

$$P(A_1|I) = P(A_2|I) = \dots = P(A_n|I) = \frac{1}{n},$$

allora preso un sottoinsieme $J \subseteq \{1, 2, \dots, n\}$,

$$P(\text{vale una tra le } A_j, j \in J|I) = \frac{\#J}{n} = \frac{\text{“casi favorevoli”}}{\text{“casi possibili”}}.$$

- Se la probabilità di tutte le alternative è la stessa,

$$P(A_1|I) = P(A_2|I) = \dots = P(A_n|I) = \frac{1}{n},$$

allora preso un sottoinsieme $J \subseteq \{1, 2, \dots, n\}$,

$$P(\text{vale una tra le } A_j, j \in J|I) = \frac{\#J}{n} = \frac{\text{“casi favorevoli”}}{\text{“casi possibili”}}.$$

- Appena si raccoglie informazione però ci si allontana da questo caso “ideale”, le probabilità si aggiornano seguendo formula di Bayes.

- **Problema:** Come **stimare** una delle alternative ad esempio per comportarsi come se sia quella effettivamente vera?

- **Problema:** Come **stimare** una delle alternative ad esempio per comportarsi come se sia quella effettivamente vera?
- Un possibile criterio è quello della **moda**, ossia dell'alternativa più probabile

$$P(A_{\text{moda}}|I) = \max \{P(A_1|I), P(A_2|I), \dots, P(A_n|I)\}.$$

- **Problema:** Come **stimare** una delle alternative ad esempio per comportarsi come se sia quella effettivamente vera?
- Un possibile criterio è quello della **moda**, ossia dell'alternativa più probabile

$$P(A_{\text{moda}}|I) = \max \{P(A_1|I), P(A_2|I), \dots, P(A_n|I)\}.$$

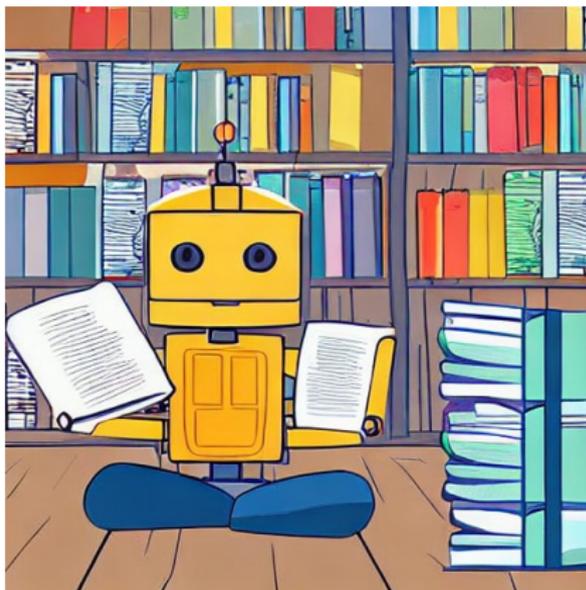
- La **teoria delle decisioni** permette di tenere conto degli eventuali **danni** derivanti da una decisione.
- La stima deve minimizzare il **rischio**, definito come

$$\text{Rischio}(i) = P(A_i|I) \cdot \text{Danno}(A_i \text{ e } I).$$

Apprendimento automatico (Machine Learning)

- I computer sono delle macchine che svolgono una grande quantità di **operazioni** seguendo precisamente le istruzioni di un **programma**.
- Tutte le operazioni (anche quelle che riguardano numeri, immagini, suoni ecc.) sono codificate e ridotte a quelle **logiche** elementari.





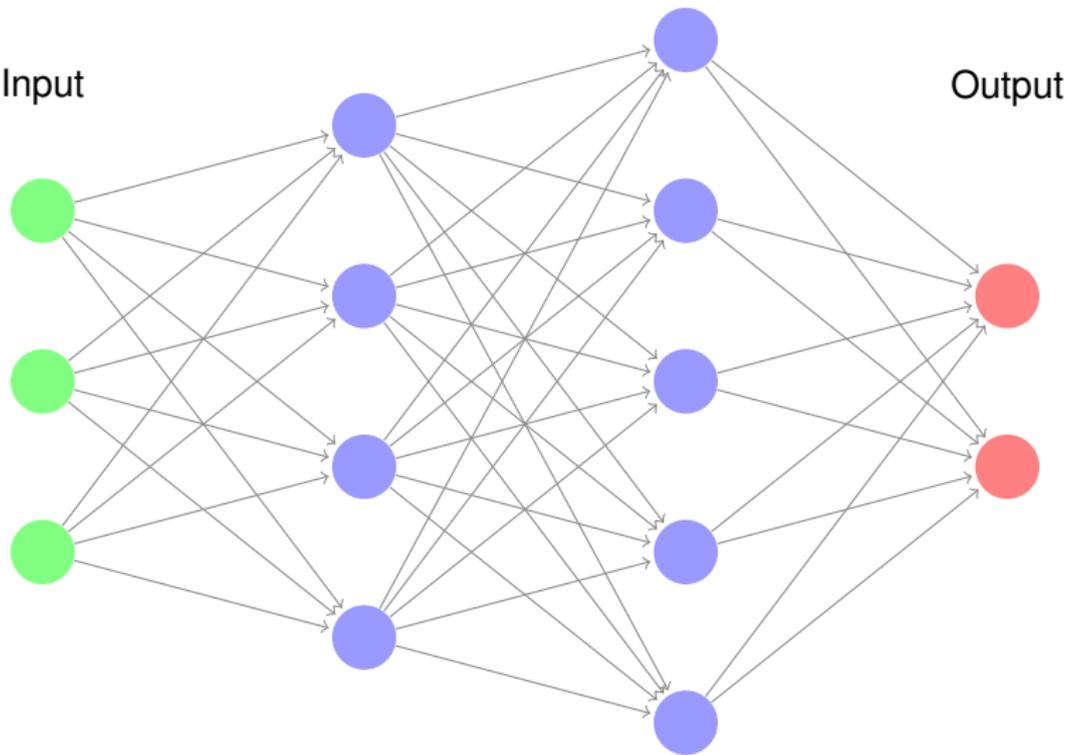
- Negli ultimi 15 anni stiamo assistendo ad una vera e propria rivoluzione, in cui la macchina **impara da sé** (o quasi) il programma.
- Questo è **tecnicamente** possibile grazie all'abbondanza di dati e alla potenza di calcolo.
- Ma il principio alla base è quello dell'**inferenza**.

Reti neurali



- Come si realizza una macchina che **impara**?
- L'idea è di imitare (ma non troppo) la biologia del cervello, costituito da **neuroni** connessi tra loro.
- Le **reti neurali profonde** (Deep Neural Networks) sono il modello matematico (più diffuso)

Input



Output

- Ciascun **neurone** contiene un valore numerico (l'analogo di un potenziale di membrana).
- Il livello di **input** contiene i dati da inserire nel programma: ad esempio, un testo, un'immagine o un video codificato tramite stringhe binarie.
- Il livello di **output** contiene ciò che la rete darà in uscita (ad esempio, un titolo per il testo o un'immagine associata al testo di input, ...)
- I livelli intermedi (**nascosti**) permettono alla rete di elaborare l'informazione.
- Ogni connessione tra neuroni è **pesata** (per semplificare, pensate a dei pesi che possono essere $+1$ o -1) e contribuisce al calcolo dei valori nei neuroni del livello successivo, dall'input fino all'output.

Apprendimento come inferenza

- L'apprendimento (training) consiste nell'individuare una configurazione dei pesi delle connessioni sulla base di un insieme di esempi.

Apprendimento come inferenza

- L'apprendimento (training) consiste nell'individuare una configurazione dei pesi delle connessioni sulla base di un insieme di esempi.
- Pensiamo alle configurazioni possibili come un sistema di alternative

$$A_1, A_2, A_3, \dots, A_n$$

- A priori, diamo la stessa probabilità a ciascuna configurazione.

Apprendimento come inferenza

- L'**apprendimento** (training) consiste nell'individuare una configurazione dei pesi delle connessioni sulla base di un insieme di esempi.
- Pensiamo alle configurazioni possibili come un sistema di alternative

$$A_1, A_2, A_3, \dots, A_n$$

- A priori, diamo la stessa probabilità a ciascuna configurazione.
- Supponiamo di avere raccolto un insieme \mathcal{T} di **input-output** che vogliamo "insegnare" alla rete (detto training set).

Apprendimento come inferenza

- L'**apprendimento** (training) consiste nell'individuare una configurazione dei pesi delle connessioni sulla base di un insieme di esempi.
- Pensiamo alle configurazioni possibili come un sistema di alternative

$$A_1, A_2, A_3, \dots, A_n$$

- A priori, diamo la stessa probabilità a ciascuna configurazione.
- Supponiamo di avere raccolto un insieme \mathcal{T} di **input-output** che vogliamo "insegnare" alla rete (detto training set).
- Per ciascuna rete A_i , calcoliamo la probabilità a posteriori

$$P(A_i|\mathcal{T}) = \frac{1}{n} \cdot \frac{L(A_i; \mathcal{T})}{P(\mathcal{T})}.$$

Apprendimento come inferenza

- L'**apprendimento** (training) consiste nell'individuare una configurazione dei pesi delle connessioni sulla base di un insieme di esempi.
- Pensiamo alle configurazioni possibili come un sistema di alternative

$$A_1, A_2, A_3, \dots, A_n$$

- A priori, diamo la stessa probabilità a ciascuna configurazione.
- Supponiamo di avere raccolto un insieme \mathcal{T} di **input-output** che vogliamo "insegnare" alla rete (detto training set).
- Per ciascuna rete A_i , calcoliamo la probabilità a posteriori

$$P(A_i|\mathcal{T}) = \frac{1}{n} \cdot \frac{L(A_i; \mathcal{T})}{P(\mathcal{T})}.$$

- Selezioniamo infine una configurazione in cui la probabilità è massima.

Dalla teoria alla pratica



- Il numero di connessioni nelle reti attualmente all'avanguardia è di centinaia di miliardi (10^{11}).
- Quindi il numero di **possibili configurazioni** è dell'ordine di $2^{10^{11}}$.

Dalla teoria alla pratica



- Il numero di connessioni nelle reti attualmente all'avanguardia è di centinaia di miliardi (10^{11}).
- Quindi il numero di **possibili configurazioni** è dell'ordine di $2^{10^{11}}$.
- Non basterebbero tutti gli atomi dell'universo per elencarle tutte!
- Per trovare la configurazione migliore, si usano quindi **metodi numerici** di ottimizzazione, ma questa è un'altra storia. . .

Conclusione

- A partire da semplici regole, il calcolo delle probabilità può portare molto lontano (sia nella matematica astratta, sia nelle applicazioni!)
- Se lo incontrerete nel vostro percorso scolastico, ricordatevi che non è utile (solo) per il gioco d'azzardo!

Conclusione

- A partire da semplici regole, il calcolo delle probabilità può portare molto lontano (sia nella matematica astratta, sia nelle applicazioni!)
- Se lo incontrerete nel vostro percorso scolastico, ricordatevi che non è utile (solo) per il gioco d'azzardo!
- Letture consigliate:
 - *Il segnale e il rumore. Arte e scienza della previsione* di Nate Silver
 - *Probability Theory: The Logic of Science* di E.T. Jaynes (primi capitoli)
 - *Saggio filosofico sulle probabilità* di P.S. Laplace (storico)

Conclusione

- A partire da semplici regole, il calcolo delle probabilità può portare molto lontano (sia nella matematica astratta, sia nelle applicazioni!)
- Se lo incontrerete nel vostro percorso scolastico, ricordatevi che non è utile (solo) per il gioco d'azzardo!
- Letture consigliate:
 - *Il segnale e il rumore. Arte e scienza della previsione* di Nate Silver
 - *Probability Theory: The Logic of Science* di E.T. Jaynes (primi capitoli)
 - *Saggio filosofico sulle probabilità* di P.S. Laplace (storico)
- Le immagini delle slides sono ottenute tutte tramite **modelli generativi**:
 - **DALL-E 2**, `labs.openai.com`
 - **Stable Diffusion**, `stablediffusionweb.com`, `lexica.art`
 - **Craiyon**, `craiyon.com`
- Provate anche i modelli linguistici: `chat.openai.com`

Input: *scrivi una canzone rap sulla probabilità, l'inferenza, la formula di Bayes e le reti neurali*

Intro Listen up, it's time to learn about probability
Inference, Bayes' Formula, and neural
networks, you see
Gonna break it down, make it simple and
plain
So you understand the power of these
scientific domains

Verse Probability, it's all about the odds
Of an event happening, it's the chance it
nods
In the right direction, giving us a clue
About what might happen, what's likely to
ensue

Chorus *Bayes' Formula, it's the key
To unlock the power of probability
Inference and networks, they all come
together
To help us understand the world in better
weather*

V2 Inferencing, it's making a guess
Based on the data, it's putting it to the test
Using Bayes' Formula, we can improve our
guess
With the power of probability, we won't
miss

Repeat Chorus

Bridge Neural networks, they're like the brain
Using algorithms, they learn and train
To make predictions, based on the data
And with Bayes' Formula, they won't
hesitate

Repeat Chorus

Outro So don't be afraid of probability
Inference, Bayes' Formula, and neural
networks, they're all family
Working together, to help us understand
The world around us, and make a better
plan.