Causal Networks

Giuseppe Magazzù

2021 - 2022

Contents

Chapter 1

The Potential Outcome Framework

Fundamental Problem of Causal Inference:

- Average Treatment Effects and Missing Data Interpretation
- Ignorability - Exchangeability
- Conditional Exchangeability - Uncounfoundedness
- Positivity - Overlap - Common Support and Extrapolation
- No interference, Consistency, SUTVA

1.1 Potential Outcome

- X **Treatment** variabile aleatoria
- Y **Outcome** variabile aleatoria
- Z **Covariate** insieme di variabili aleatorie

Il **potential outcome** $Y(x)$ denota quale sarebbe l'**outcome** quando $X = x$, ovvero se il treatment che è stato scelto è x. Una volta che viene osservato il **potential outcome** Y(x) questo assume valore Y chiamato **outcome**.

Una **popolazione** consiste di molti individui o unità. Ogni individuo (unità) è associato a uno o più variabili Z **covariate**.

Denotiamo X, Y, Z dell'individuo i-esimo come X_i, Y_i, Z_i.

$Y(x)$ è una variabile aleatoria
$Y_i(x)$ non è trattata come una variabile aleatoria poiché specifica per l'individuo

Individual Treatment Effect (ITE)

Per verificare se c'è una relazione causale tra X e Y si può calcolare la seguente differenza $\tau_i \triangleq Y_i(1) - Y_i(0)$.

Se il potential outcome è uguale in entrambi i casi ($\tau_i = 0$) allora non c'è relazione causale. Nel caso contrario si può notare che diversi treatment portano a diversi outcome e quindi c'è una relazione causale.

Non possiamo osservare tutti i **potential outcome** poiché l'osservarne uno influenzerebbe gli altri. I **potential outcome** che non possono essere osservati vengono chiamati **counterfactual**, mentre quello che osserviamo è il **factual**.

$$
\begin{array}{lllll}
Y(1) = ? & \Rightarrow & Y = 0, X = 1 & \Rightarrow & Y(1) = ? \quad \text{counterfactual} \\
Y(0) = ? & \Rightarrow & Y = 0, X = 0 & \Rightarrow & Y(0) = 1 \quad \text{factual}
\end{array}
$$

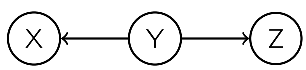
Average Treatment Effect (ATE)

$\tau \triangleq \mathbb{E}[Y_i(1) - Y_i(0)]$

Chapter 2

Flow of Associations and Causal Graphs

In un grafo diretto indichiamo con $pa(X)$ i **genitori** del nodo X e $ch(Y)$ i **figli** del nodo Y.



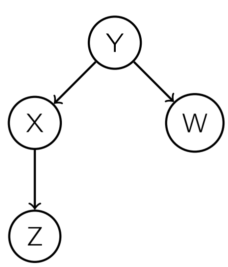
$$pa(Y) = \emptyset \quad ch(Y) = \{X, Z\}$$
$$pa(X) = Y \quad ch(X) = \emptyset$$

Un **cammino** è una qualsiasi sequenza di nodi adiacenti, indipendentemente dalla direzione degli archi che li collega.

Un **cammino diretto** è un qualsiasi **cammino** tra due nodi in cui tutti gli archi che li collega hanno la stessa direzione.

$de(Y)$ è l'insieme dei nodi **discendenti** del nodo Y, ovvero tutti i nodi che possono essere raggiunti da Y.

$an(Z)$ è l'insieme dei nodi **antenati** del nodo Z, ovvero tutti i nodi che da Z possono essere raggiunti.



$$an(Z) = \{X, Y\}$$
$$an(Y) = \emptyset$$

$$de(Y) = \{X, Z, W\}$$
$$de(W) = \emptyset$$

2.1 Bayesian Networks

In una rete bayesiana vogliamo modellare la distribuzione di probabilità $P(X_1, \ldots, X_n)$. Tramite la **Chain Rule** possiamo riscriverla nel seguente modo:

$$
\begin{aligned}
P(X_1, \ldots, X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\cdots P(X_n|X_1, \ldots, X_{n-1}) \\
&= P(X_1)\prod_{i=2}^{n} P(X_i|X_1, \ldots, X_{i-1})
\end{aligned}
$$

Local Markov Assumption: Un nodo X è indipendente da tutti i suoi non-discendenti date l'evidenze di tutti i nodi genitori $pa(X)$.

Data questa assunzione possiamo applicare la fattorizzazione della probabilità P.

$$
P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|pa(X_i))
$$

In questo modo possiamo ridurre il numero di parametri della rete.

Una distribuzione di probabilità P si dice che sia Markov se ogni nodo X rispetta la Local Markov Assumption.

L'assunzione di Markov non ci da informazioni riguardo relazioni di dipendenza tra i nodi. Quindi estendiamo quest'assunzione.

Minimality Assumption

- Dato $pa(X)$, un nodo X è indipendente da tutti i suoi non-discendenti.
- I nodi adiacenti sono dipendenti.

Dato un DAG G, se P è Markov allora sappiamo:

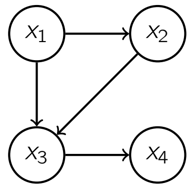
- P soddisfa un insieme di indipendenze specificate dalla struttura di G.
- se P soddisfa pure la Minimality Assumption, allora l'insieme di indipendenze è minimale, ovvero P non soddisfa altre indipendenze in G. Questo equivale a dire che tutti i nodi adiacenti sono dipendenti.

2.2 Causal Networks

Una variabile X si dice **causa** di una variabile Y se Y può cambiare in risposta a un cambiamento di X.

Causal Edge Assumption: Ogni variabile associata a un nodo è causata dalle variabili dei nodi genitori.

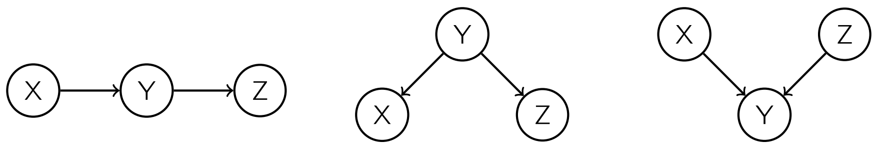
Un **Grafo Causale** è un DAG in cui è soddisfatta la proprietà **Causal Edge Assumption**.



X_1 causa direttamente X_2 e X_3
X_2 causa direttamente X_3
X_3 causa direttamente X_4

X_1 causa indirettamente X_4

Per capire la differenza tra association flow e causation flow utilizziamo dei grafi elementari.



(a) Chain

(b) Fork

(c) Collider

Il flusso di associazione esprime che due nodi del grafo sono associati o meno.

Vogliamo sapere se due nodi sono (statisticamente) dipendenti o indipendenti.

Un-connected Nodes

Due nodi sconnessi non sono associati. Questo può essere dimostrato applicando la fattorizzazione.

$$
\begin{aligned}
P(X,Y) &= P(X|Y)P(Y) && \text{(probabilità composta)} \\
&= P(X|pa(X))P(Y|pa(Y)) && \text{(fattorizzazione)} \\
&= P(X)P(Y) && \text{indipendenza}
\end{aligned}
$$

Due nodi connessi adiacenti sono dipendenti per la **local edge assumption**.

Chain e Fork

Collider

2.3 D-Separation

Un percorso p si dice bloccato da un insieme di nodi S sse

- p contiene una chain $A \to B \to C$ o una fork $A \leftarrow B \to C$ e $B \in S$
- p contiene un collider $A \to B \leftarrow C$ e $B \notin S, de(B) \nsubseteq S$

Se l'insieme S blocca ogni percorso tra due nodi X e Y, allora X e Y si dicono **d-separati** da S, e quindi indipendenti dato S.

ESEMPI...

La d-separazione implica l'indipendenza condizionale.

- Due variabili X e Y sono d-separate nel grafo G quando condizionate dall'insieme S ($X \perp\!\!\!\perp_G Y|S$).
- Due variabili X e Y sono indipendenti nella distribuzione p quando condizionate dall'insieme S ($X \perp\!\!\!\perp_p Y|S$).

$X \perp\!\!\!\perp_G Y|S \Rightarrow X \perp\!\!\!\perp_p Y|S$

Chapter 3

Causal Models

Chapter 4

Structural Causal Models

Chapter 5

Randomized Experiments

Chapter 6

Nonparametric Identification

Chapter 7

Estimation

Chapter 8

Unobserved Confounding

Chapter 9

Instrumental Variables

Chapter 10

Causal Discovery from Observational Data

Chapter 11

Transfer Learning and Transportability

Chapter 12

Counterfactuals