

Causal Reasoning in Machine Learning

Pier Paolo Ippolito



Benvenuti!

- SAS Institute Data Scientist
- Scrittore per Towards Data Science (Tech4future in Italiano)
- Freelancer
- MSc in Artificial Intelligence (University of Southampton)



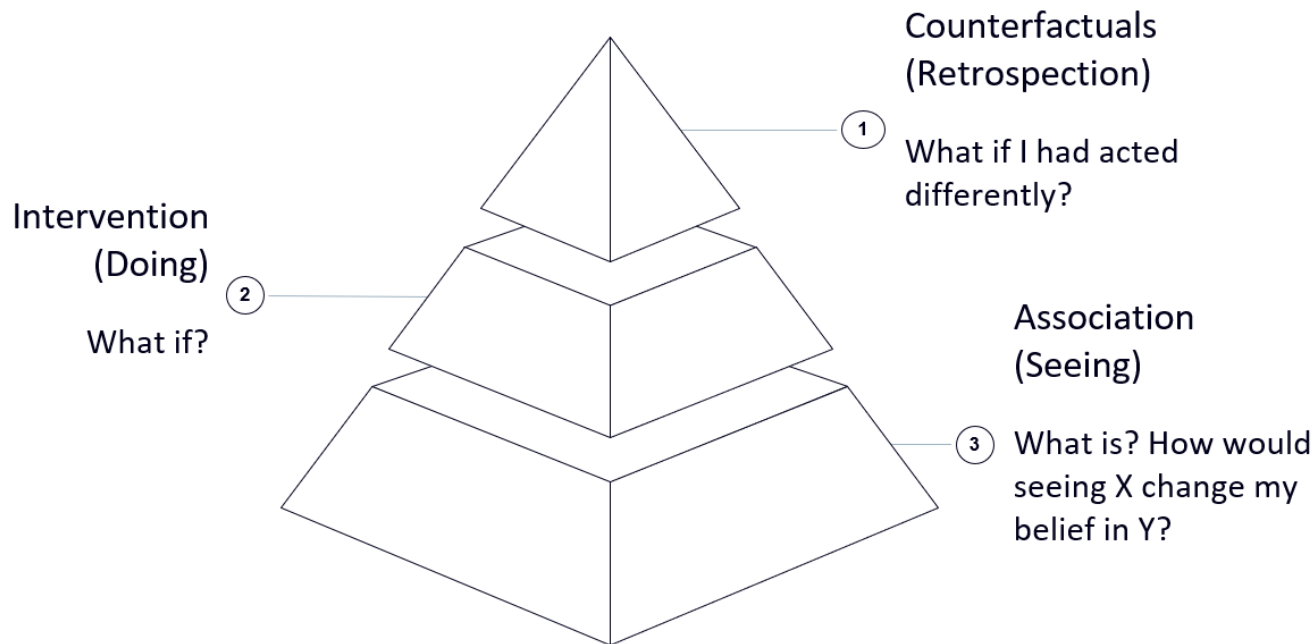
1. Causal Reasoning

Machine Learning

Il problema

- Al giorno d'oggi le tecnologie di Machine Learning si basano solo sulle correlazioni tra le diverse “features”. Ciò nonostante, questo approccio può eventualmente portare a conclusioni errate poiché correlazioni non implicano necessariamente causalità.
- Tre dei principali limiti dei modelli basati su Machine Learning e Deep Learning sono: Robustezza, Spiegabilità e Dipendenza dai dati.

Gerarchia della causalità



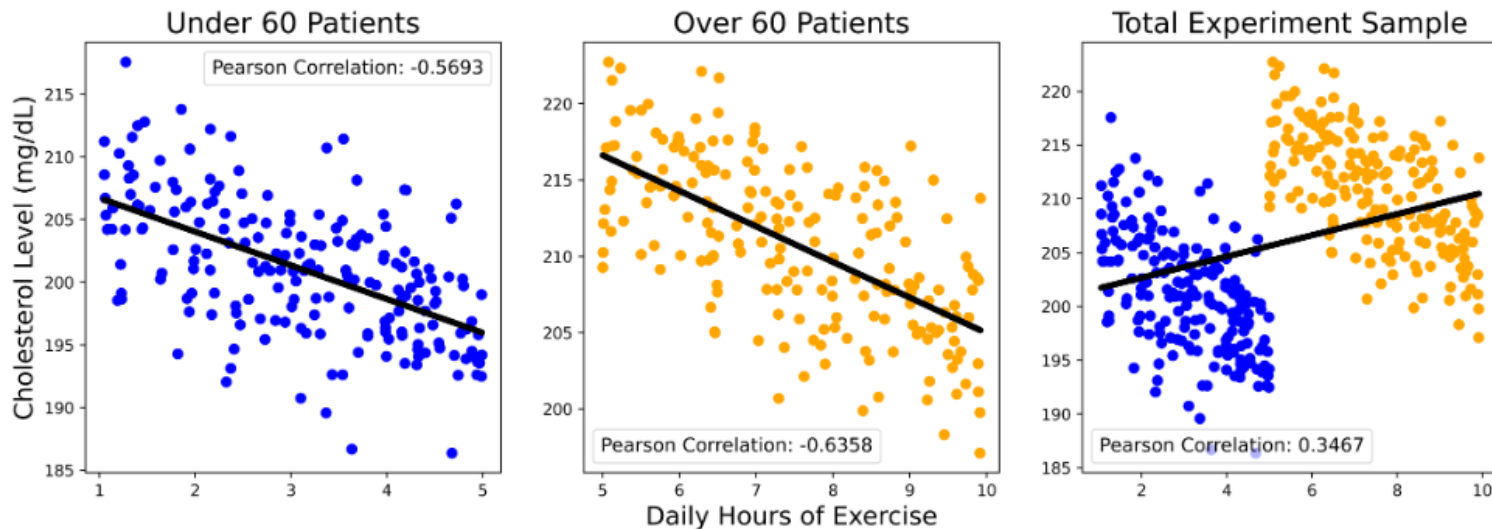
The Book of Why:
The New Science
of Cause and Effect
(Judea Pearl)

Il paradosso di Simpson

«In statistica, il paradosso di Simpson indica una situazione in cui una relazione tra due fenomeni appare modificata, o perfino invertita, dai dati in possesso a causa di altri fenomeni non presi in considerazione nell'analisi (variabili nascoste)»

Il paradosso di Simpson

Cholesterol vs Daily Hours of Exercise



Il paradosso di Simpson





2.

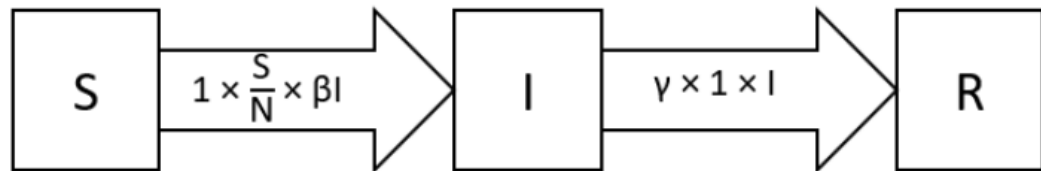
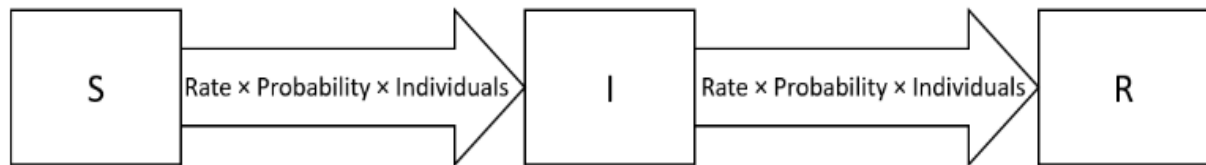
Modelli e Simulazioni

Data Science senza dati

“ Tutti i modelli sono sbagliati, ma alcuni sono utili.

George E. P. Box, Statistics for Experimenters, second edition, 2005, page 440

Modelli Compartmentali



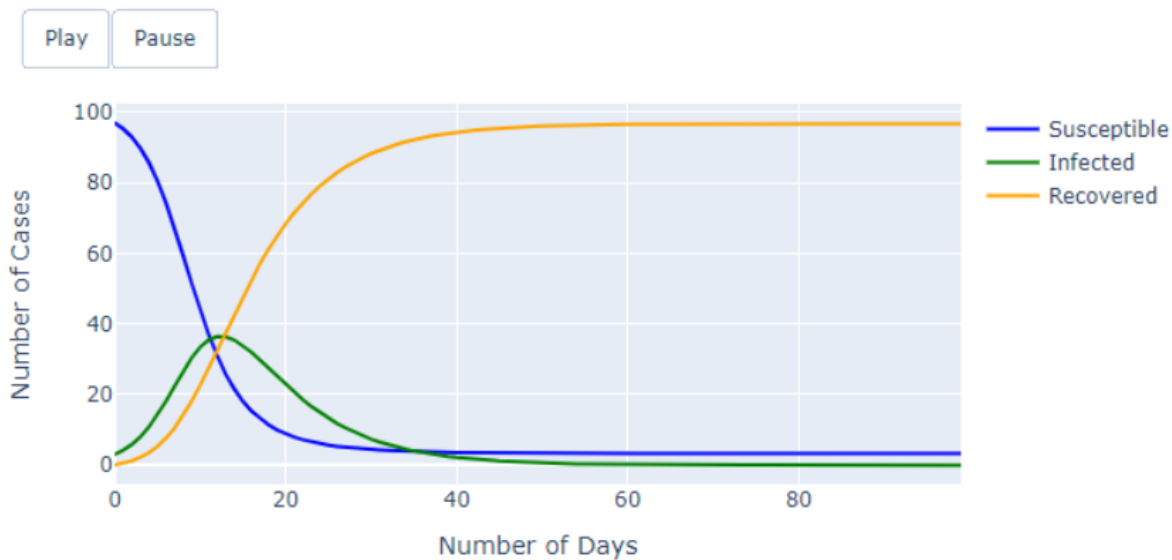
$$\frac{\partial S}{\partial t} = -\beta \times I \times \frac{S}{N}$$

$$\frac{\partial I}{\partial t} = \beta \times I \times \frac{S}{N} - \gamma \times I$$

$$\frac{\partial R}{\partial t} = \gamma \times I$$

Modelli Compartmentali

Standard SIR Model ($R_0=3.5$)



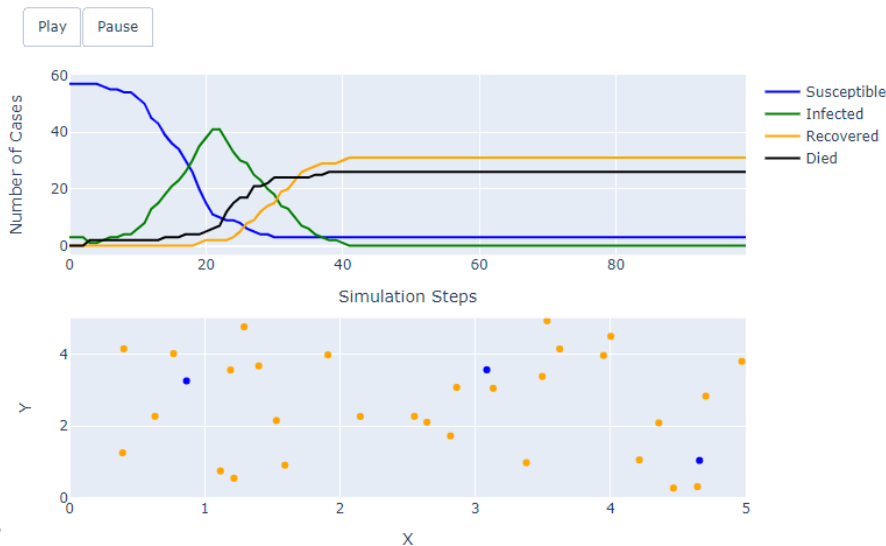
Modellazione basata su agenti (Programmazione ad oggetti)

```
class Person:
    def __init__(self, situation, position, speed, extremes, age):
        self.situation, self.position = situation, position
        self.speed, self.extremes = speed, extremes
        angle = np.random.uniform(0, 2*np.pi)
        self.x_dir, self.y_dir = np.cos(angle), np.sin(angle)
        self.age = age
        self.rehabilitation = 0

    def step_ahead(self, community, proximity, contagiousness,
                  no_move, deaths_p):
        self = contacts(self, community, proximity, contagiousness,
                        deaths_p)
        self = new_pos(self, no_move)
```

Modellazione basata su agenti

Interactive Simulation Modelling



Algorithm 1 Population Modelling Pseudo-code Outline

```

1:  $E \leftarrow \text{Contact Radius}$ 
2:  $\bar{p} \leftarrow \text{Unlikeliness of Spread}$ 
3:  $p_{\text{died}} \leftarrow \text{Death Probability dependent on age}$ 
4: for day in simulation days do
5:   for individual in population do
6:     Record individual status
7:     if Infected then
8:       if Draw with probability ( $p_{\text{died}} \times \text{age} == 1$ ) then
9:         individual status  $\leftarrow$  Dead
10:      else
11:        if (Rehabilitation days == 14) then
12:          individual status  $\leftarrow$  Recovered
13:        end if
14:        Rehabilitation days  $+= 1$ 
15:      end if
16:    else if Susceptible then
17:      close_people = 0
18:      for friend in community do
19:        if (Friend == Infected) and (Euclid Dist <  $E$ ) then
20:          close_people  $+= 1$ 
21:        end if
22:      end for
23:      if (Draw with probability dependent on close_people and  $\bar{p} == 1$ ) then
24:        individual status  $\leftarrow$  Infected
25:      end if
26:    end if
27:    if (Static == False) then
28:      individual X and Y position update
29:      if individual X or Y position out of boundaries then
30:        Adjust position and reverse movement direction
31:      end if
32:    end if
33:  end for
34: end for

```

3.

Studio Epidemiologico

COVID-19

COVID-19



Applicazione Web

- ✓ Distribuzione su AWS (Amazon Web Services) EC2 Linux
- ✓ Interfaccia Streamlit
- ✓ Supporto per Docker
- ✓ Dati riguardo COVID-19 aggiornati in automatica ogni 24 ore
- ✓ News internazionali aggiornate ogni 2 ore

Applicazione Web

The screenshot shows the GitHub interface for the repository 'pierpaolo28 / Epidemics-Modelling'. The repository is private and has 4 branches and 0 tags. The main branch is 'master'. The repository description is 'Modelling and tracking of epidemics data analysis in Python and Julia'. The repository has 3,22,240,181,850 stars and 1 fork. The repository is licensed under MIT License. The repository has no releases published and no packages published. The repository has 1 environment, 'github-pages', which is active. The repository has 1 language, 'Python', which is 0.4% of the code.

Repository: pierpaolo28 / Epidemics-Modelling (Private)

Navigation: <> Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, Settings

Branches: master (4 branches), 0 tags

Files:

File	Description	Last Commit
.github/workflows	Pull Requests automated labelling	last month
data	Coronavirus Modelling	2 months ago
dist	updated README	3 months ago
docs	Structuring repository	3 months ago
notebooks	Cleaning Notebook	2 months ago
src	Typos fix	2 days ago
.gitattributes	Initial commit	3 months ago
.gitignore	News API	2 months ago
LICENSE	Initial commit	3 months ago
README.md	Update README.md	3 months ago
requirements.txt	Updated requirements	last month

README.md

forks [repo not found](#) stars [repo not found](#) issues [repo not found](#) [LinkedIn](#)

Epidemics Modelling

Table of Contents

- [About the Project](#)
- [Getting Started](#)

About

Modelling and tracking of epidemics data analysis in Python and Julia

[3.22.240.181.850](#)

[epidemiology](#) [covid-19](#) [python](#)

[julia](#) [dashboard](#)

Readme

MIT License

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Environments 1

[github-pages](#) Active

Languages

Jupyter Notebook 99.6%
Python 0.4%

Dimostrazione Pratica

- ✓ Pubblicazione Completa
- ✓ Extras
- ✓ Open Source Codice
- ✓ Applicazione Web

A background image showing a white notebook with horizontal lines and two pens. One pen is pink with a black grid pattern, and the other is white with gold dots. The image is partially covered by a white diagonal shape.

Grazie!

Domande?

Contatti:

- [LinkedIn](#)
- [GitHub](#)
- [Online Portfolio](#)
- [Towards Data Science](#)