

Lecture 12: Machine Learning in the Real World

Xenia Miscouridou

Imperial College London

Machine Learning Summer School
July 2022

1 Recap

2 Real world applications of machine learning

3 Responsible machine learning rules

Outline

- 1 Recap
- 2 Real world applications of machine learning
- 3 Responsible machine learning rules

Let's revisit the course objectives

- ① To provide an overview of machine learning
- ② To describe some common machine learning approaches for supervised and unsupervised learning techniques
- ③ To explain how these algorithms are trained given some data
- ④ To discuss the concept of generalisation
- ⑤ To learn how to apply supervised and unsupervised learning algorithms to real data in R

What have we missed?

- Bayesian methods and models
- Probabilistic graphical models
- Causal inference
- Computational statistical methods, especially MCMC and approximate Bayesian inference
- Network Modelling

What have we missed?

- Bayesian methods and models
- Probabilistic graphical models
- Causal inference
- Computational statistical methods, especially MCMC and approximate Bayesian inference
- Network Modelling

What have we missed?

- Bayesian methods and models
- Probabilistic graphical models
- Causal inference
- Computational statistical methods, especially MCMC and approximate Bayesian inference
- Network Modelling

What have we missed?

- Bayesian methods and models
- Probabilistic graphical models
- Causal inference
- Computational statistical methods, especially MCMC and approximate Bayesian inference
- Network Modelling

What have we missed?

- Bayesian methods and models
- Probabilistic graphical models
- Causal inference
- Computational statistical methods, especially MCMC and approximate Bayesian inference
- Network Modelling

Where can I learn more?

- Machine Learning: A Probabilistic Perspective [Murphy 2012]
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction [Hastie, Tibshirani, Friedman 2009]
- Deep Learning [Courville, Goodfellow, Bengio 2015]
- A First Course in Machine Learning [Rogers and Girolami 2016]
- Information Theory, Inference, and Learning Algorithms [Mackay, 2005]
- Deep Learning with R [Chollet and Allaire 2018]

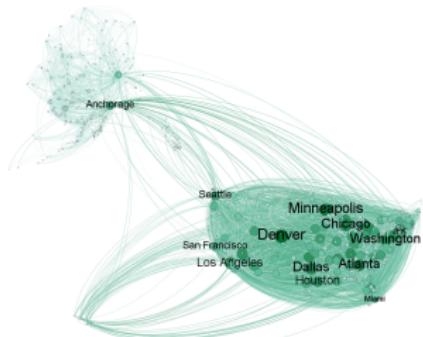
Outline

- 1 Recap
- 2 Real world applications of machine learning
- 3 Responsible machine learning rules

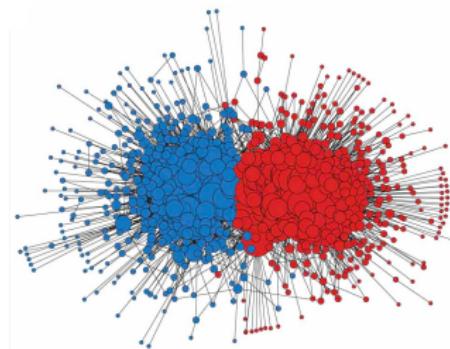
Networks from various domains



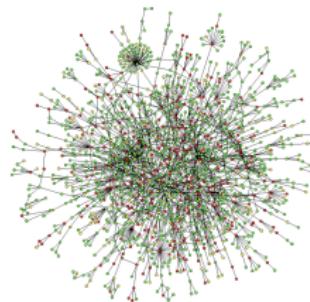
Friendships: Facebook 2010



Transportation: Airport US connections 2010



Information: US political blog citations 2005

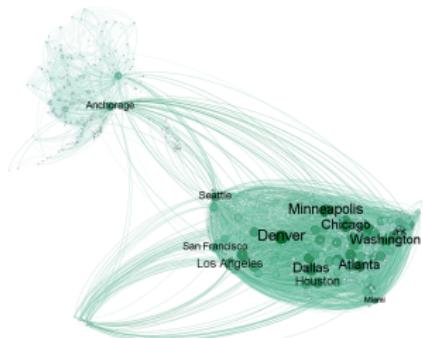


Biology: Yeast protein interaction network 2003

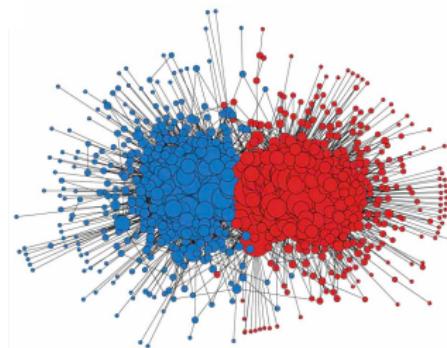
Networks from various domains



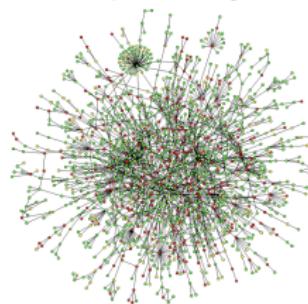
Friendships: Facebook 2010



Transportation: Airport US connections 2010



Information: US political blog citations 2005



Biology: Yeast protein interaction network 2003

Simulate a graph using our methods

Daily activity of Google maps users



- Mobility data have information about the daily activity of users
- How can I learn lower dimensional representations to cluster customers?

Identifying a new COVID-19 variant

NEWS

Home Book Coronavirus UK World Business Politics Tech Science Health Family & Education

Health

Covid-19: New variant 'raises R number by up to 0.7'

24 hours ago



New mutant coronavirus strain IS more contagious by nearly 50%

Daily Mail • 2 Jan

Covid variant spread rapidly in England even during lockdown, study finds

Government caves in to demands and closes all London primary schools in latest U-turn

Please keep your distance

Two people wearing face masks walking past a blue sign that says "Please keep your distance".

Government caves in to demands and closes all London primary schools in latest U-turn

24 hours ago

sky news

COVID-19: Brazilian P1 variant may spread more easily and could evade immune system, scientists find

Experts from Brazil and the UK estimate the variant is more transmissible than previous versions of coronavirus found in Manaus.

New mutant coronavirus strain IS more contagious by nearly 50%

Brazilian Covid variant: what do we know about P1?

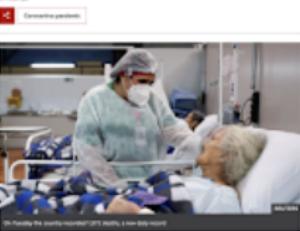
Six cases have been detected in Britain. What threat does the variant pose, and how is it different?

- Coronavirus - latest updates
- See all our coronavirus coverage



Emergency services on alert during the Covid-19 crisis in São Paulo, Brazil. The city has increased restrictions to tackle the pandemic. Photograph: Getty Images/Alamy Live News

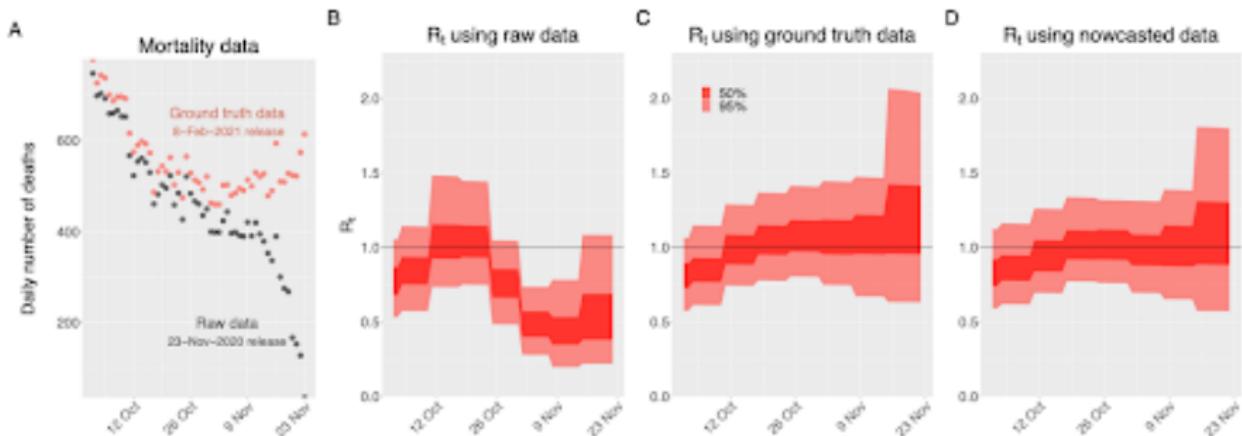
Covid-19: Brazil experts issue warning as hospitals 'close to collapse'



Meeting the country's record 10,000 daily new cases

- Characterise a new variant and its transmissibility
- Influence the measures taken in Brazil and suggested by WHO

Correcting for mortality reporting



- Current reported number of deaths is wrong
- This can lead to wrong estimates about the evolution of the pandemic
- The numbers are being updated weekly retrospectively

Outline

- 1 Recap
- 2 Real world applications of machine learning
- 3 Responsible machine learning rules

Responsible Machine Learning Principles

1. Human Augmentation
2. Bias evaluation
3. Explainability by justification
4. Reproducible Operations
5. Trust by Privacy
6. Data Risk Awareness

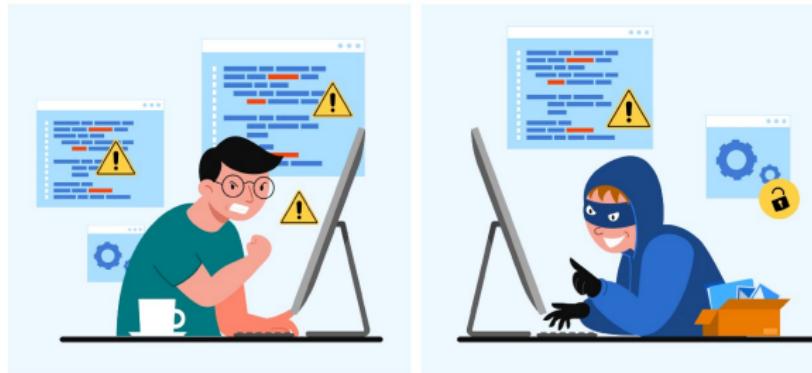
1. Human Augmentation

Assess the impact of incorrect predictions and, when reasonable, design systems with human-in-the-loop review processes

Can you think of examples where human augmentation is important?

1. Human Augmentation

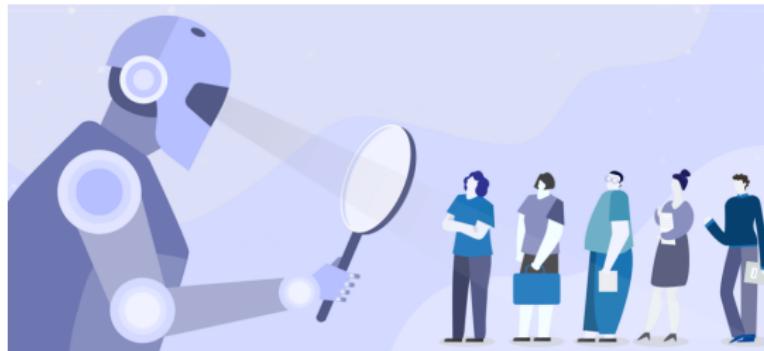
- **Fraud detection prediction** Instead of fully removing humans from the process completely, a domain expert can be requested to verify some of the results from the model to ensure the performance is aligned with the objectives.



2. Bias evaluation

Develop processes that allow me to understand, document and monitor bias in development and production

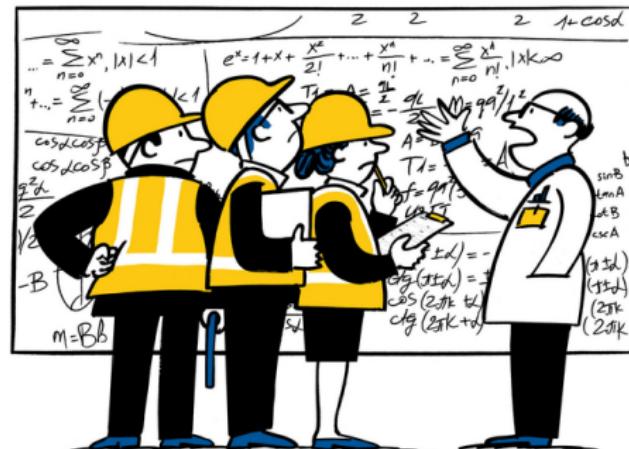
- Google's what if tool provides an interactive way to visualise and assess for model and data bias - for example in the income classification datasets it's possible to see that "race" and "gender" are two of the strongest features.
- The deployment of a biased system, can have the effect of reinforcing that pre-existing societal bias



3. Explainability by justification

Develop tools and processes to continuously improve transparency and explainability of machine learning models where reasonable

- Interpretability and explainability
 - Feature importance
 - Domain knowledge



4. Reproducible operations

Give the infrastructure required to enable for a reasonable level of reproducibility across the operations of ML systems

- Abstracting each computational step
- Adopting open standards



5. Trust by privacy

Build and communicate processes that protect and handle data with stakeholders that may interact with the system directly and/or indirectly.

- Right process and technologies are in place to protect personal data
- Personal data via metadata



6. Data risk awareness

Develop and improve reasonable processes and infrastructure to ensure data and model security are being taken into consideration during the development of machine learning systems

What are some examples where I should focus to become aware of potential risks in my data and models?

6. Data risk awareness

Develop and improve reasonable processes and infrastructure to ensure data and model security are being taken into consideration during the development of machine learning systems

- Adversarial patch tricking models
- Data breaches are caused due to simple human errors



References

- Exchangeable random measures for sparse and modular graphs with overlapping community structure, Todeschini, Mincioidou, Caron, JRSSB, 2020.
- Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil, Faria, Mellan, Mishra et al, Science, 2021
- Gaussian Process Nowcasting: Application to COVID-19 Mortality Reporting, Hawlyluk et al, UAI, 2021
- <https://ethical.institute/principles.html>
- <https://pair-code.github.io/what-if-tool/>