Review and Outlook

Machine Learning 2020
mlvu.github.io

the plan

part 1:

Review, Exam strategies

Future work
Causality, Generalization, Compositionality

part 2:

The social impact of machine learning

mind maps

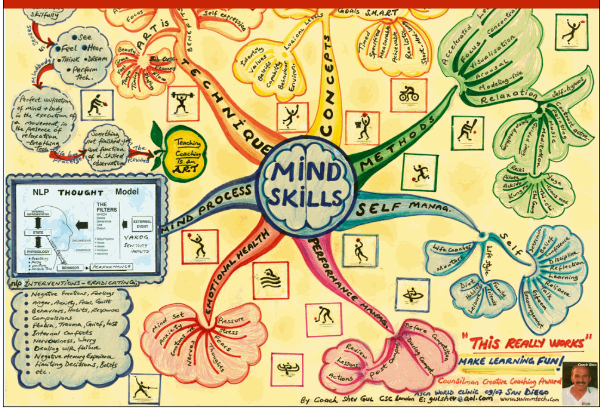


image source: Shev Gul

Mind Skills mind map. Created by Shev Gul

http://www.mindmapart.com/

exam strategy

Focus on the LECTURES, not the reading
Read the slides before watching the videos

Focus on the first 10 lectures

Make quick passes over everything. Figure out *what* you don't understand, then move on.

studying tricks

Compose a keyword list

Pages > Terminology

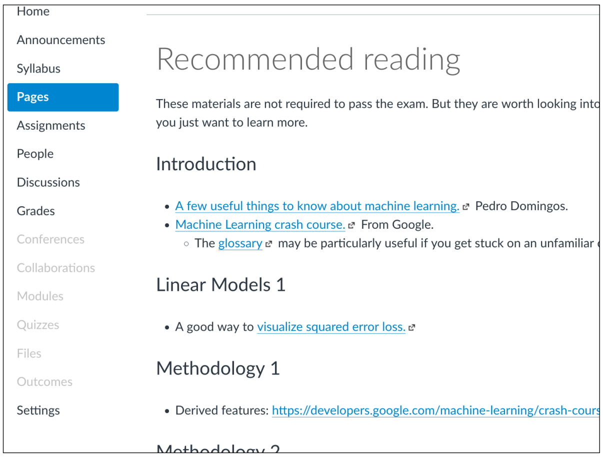
Come up with your own exam questions

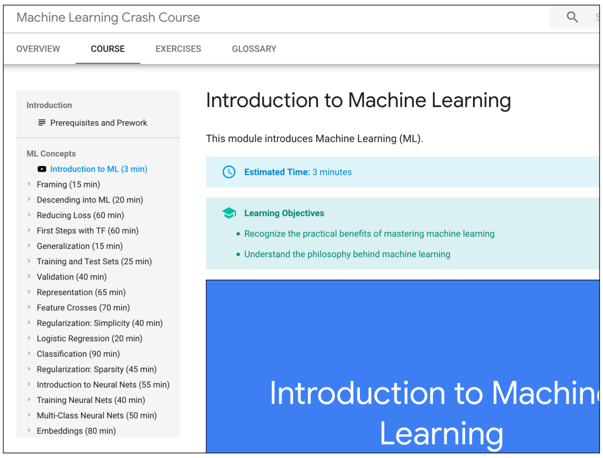
Make random combinations

focus on the ins and outs

$$
\begin{aligned}
\nabla \mathbb{E}_a r(a) &= \nabla \sum_a p(a) r(a) \\
&= \sum_a \nabla p(a) r(a) \\
&= \sum_a p(a) \frac{\nabla p(a)}{p(a)} r(a) \\
&= \sum_a p(a) \nabla \ln p(a) r(a) \\
&= \mathbb{E}_a r(a) \nabla \ln p(a)
\end{aligned}
$$

$$\nabla \ln(z) = \frac{1}{z} \nabla z$$





Machine Learning Crash Course

OVERVIEW COURSE EXERCISES **GLOSSARY**

Machine Learning Glossary

This glossary defines general machine learning terms as well as terms specific to Tens

A

A/B testing

A statistical way of comparing two (or more) techniques, typically an incumbent agains testing aims to determine not only which technique performs better but also to underst difference is statistically significant. A/B testing usually considers only two techniques measurement, but it can be applied to any finite number of techniques and measures.

accuracy

The fraction of predictions that a classification model got right. In multi-class classific is defined as follows:

Correct Predictions

ttps://developers.google.com/machine-learning/crash-course/glossary

---

Seeing Theory EN

Chapter 1

Basic Probability

This chapter is an introduction to the basic concepts of probability theory.

- Chance Events
- Expectation
- Variance

2018 - P 4

- Home
- Announcements
- Syllabus
- Pages
- Assignments
- People
- **Discussions**
- Grades
- Files
- Quizzes
- Outcomes
- Collaborations
- Conferences
- Modules
- Settings

All | Search by title or author...

Pinned discussions

Terminology and notation
All sections
Last post at 24 Feb at 15:59

Typos and other small mistakes
All sections
Last post at 7 Mar at 11:31

Discussions

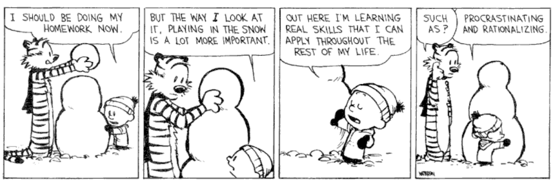
Questions exam 2018
All sections
Last post at 20 Mar at 17:04

Hw6, Decision Trees question 3
All sections
Last post at 18 Mar at 16:26

Project baseline requirements

---

break



what can't we do yet?

Causality

Compositionality

Generalization

causality

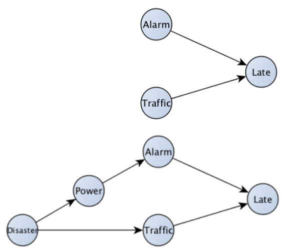
correlation does not imply causation

offline learning can only find correlations.

identifying causation requires intervention
i.e. a controlled experiment

causality without experiments

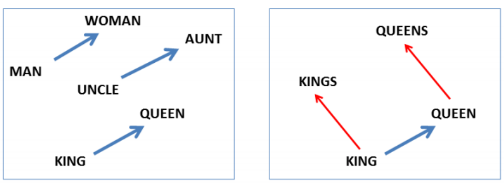
background knowledge



source:https://medium.com/causal-data-science/if-correlation-doesnt-imply-causation-then-what-does-c74f20d26438

compositionally

$$\underbrace{v(\text{king}) + v(\text{woman}) - v(\text{man})}_{\text{"feminine" vector}} \approx v(\text{queen})$$



(Mikolov et al., NAACL HLT, 2013)

generalisation

What if your test data is a little different from your training data.

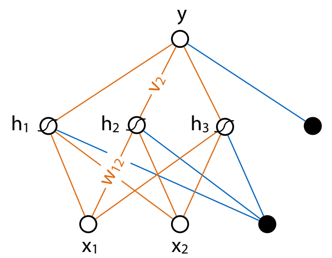
For instance:

train an RNN to sum numbers between 1 and 10
test on numbers between 1 and 15

causality, compositionality, generalisation

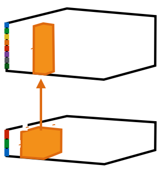
The key is to create a model with the right **inductive bias**

inductive biases: MLP



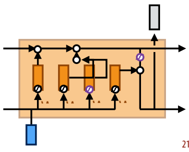
inductive biases of CNNs

- The data has a grid structure
  we know the data consists of pixels
- Inputs far apart on the grid are not relevant for low-level features
  we connect only a local group of pixels to each hidden node
- low level feature extractors are translation invariant
  we re-use the same weights for each patch



inductive biases of LSTMs

- The data is a sequence
- Each token can be modelled as a result of the tokens preceding it.
- Many tokens can be forgotten, and we can infer this from the token itself, together with the immediate context.



inductive bias

causality: inject background knowledge as an inductive bias

compositionality: add preference for compositionally explicitly, or model the rules of composition

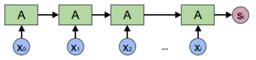
generalization: the more we constrain our model, the better it generalizes
but the less robust it is against the thing we didn't model

Grand challenge: start with the inductive bias, and let the model follow.

end-to-end learning

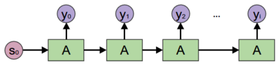
OCR → NER → REX → Rea

software 2.0 / differential programming



fold = Encoding RNN

Haskell: `foldl a s`



unfold = Generating RNN

Haskell: `unfoldr a s`

medium.com/@karpathy/software-2-0

colah.github.io/posts/2015-09-NN-Types-FP/

Using generative models to invent meaningful creative operations

Let's look at an example where a machine learning model makes a new type of interface possible. To understand the interface, imagine you're a type designer, working on creating a new font [1]. After sketching some initial designs, you wish to experiment with bold, italic, and condensed variations. Let's examine a tool to generate and explore such variations, from any initial design. For reasons that will soon be explained the quality of results is quite crude; please bear with us.

distill.pub/2017/aia/

FACES

Smiling

Images from *Sampling Generative Networks* by White [6].

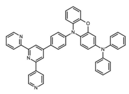
SENTENCES

"It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife."

Length

Sentence from *Pride and Prejudice* by Jane Austen. Interpolated by the authors. Inspired by experiments done by the novelist Robin Sloan [7]

MOLECULES



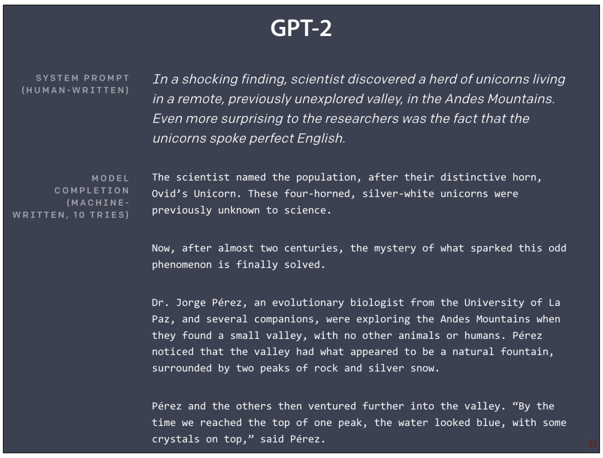
Delayed fluorescence decay rate

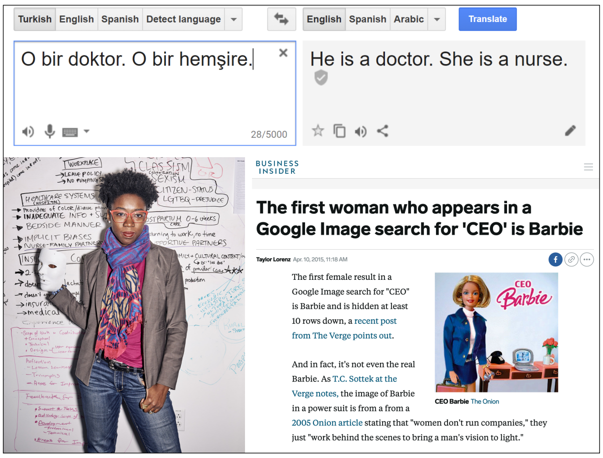
Images from *Automatic chemical design using a data-driven continuous representation of molecules* by Gómez-Bombarelli *et al* [8].

AI
artificial intelligence

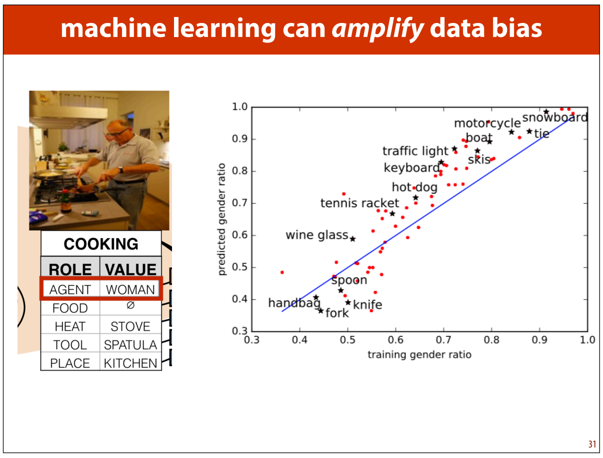
IA
intelligence augmentation

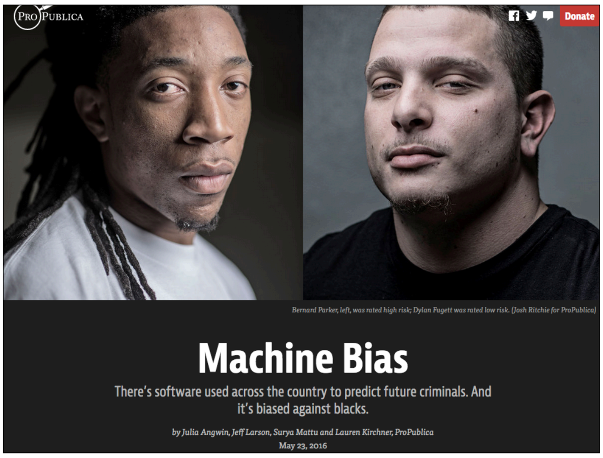
II
Intelligent Infrastructure





Joy Buolamwini





https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?utm_campaign=comms&utm_source=comms-pitch&utm_medium=email&utm_term=algorithm

---

What if

the data is a fair representation of the population

and

the predictions are accurate?

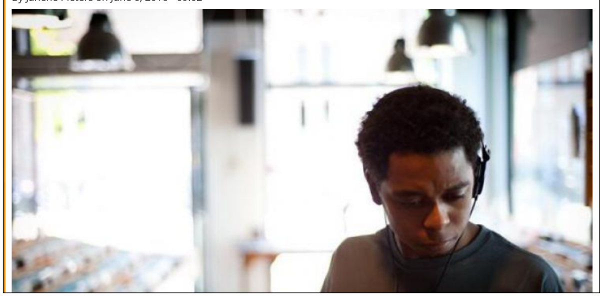
---

racial profiling

TOP STORIES

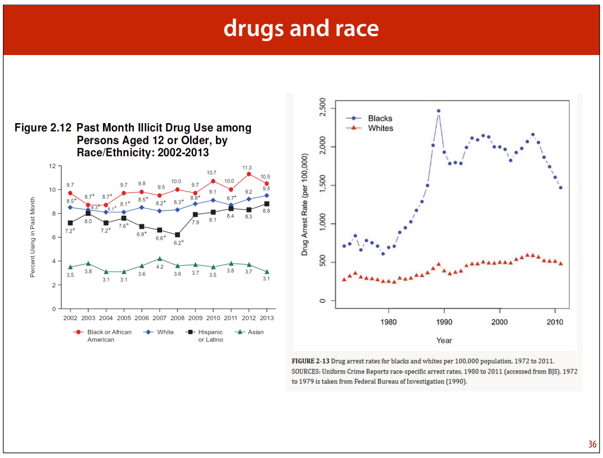
POLICE RACIAL PROFILING OVERWHELMINGLY APPROVED BY DUTCH PUBLIC

By Janene Pieters on June 6, 2016 - 09:02



Recently, a Dutch hip-hop artist called Typhoon was stopped by the police. The police admitted that the combination of his skin colour and the fact that he drove an expensive car played a part in the choice to stop him. This caused a small stir in the Dutch media and a nationwide discussion about **racial profiling**, using racial features to predict the likelihood of a person committing a crime. Other examples of profiling include travel security checking people of arabic descent more than others, giving people of certain background higher health-insurance premiums or managers being less likely to hire women for technical positions (gender profiling).

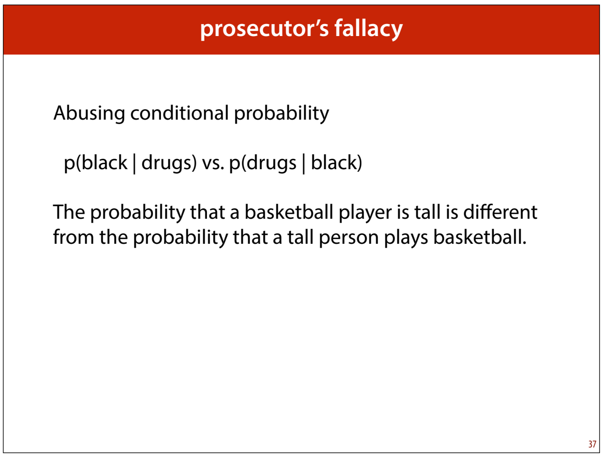
Profiling is an important subject now that machine learning and data mining are becoming more widespread. Since we generally optimise purely for performance, and feed the algorithm lots of features, there is no telling whether it is using sensitive feature like race. An automatic system built to detect whether cars should be stopped for a random search might be very effective at predicting crimes, but unless it's 100% effective, it will stop innocent people to, and it may be engaging in racial profiling.



Talking about things like "the probability that a black person commits a crime" is a reductive way to speak, regardless of our intentions, so we will try to make things more concrete with an example: we will look at this recent case, discussed in the Washington Post. In the US, people are routinely classified by ethnicity in research (since it is an important issue) and there are clear guidelines for making the classification. We will follow this definition of a black person. Instead of crime in general, we will focus on the illicit use of drugs. For this, we have good data on how many people engage in illicit drug use and for how many people are arrested for it.

Here we see the rates of illicit drug use broken down by ethnicity, and the arrest rate also broken down by ethnicity. As we see, there is a very small discrepancy between the black/white difference in illicit drug use, ad a huge margin in the difference in arrests.

source:
http://skeptics.stackexchange.com/questions/36797/do-black-people-and-white-people-use-drugs-at-the-same-rate-in-the-usa-but-blac
https://www.washingtonpost.com/news/wonk/wp/2013/06/04/the-blackwhite-marijuana-arrest-gap-in-nine-charts/?utm_term=.322fc255f412



Racial profiling is a classic case of the prosecutor's fallacy. In this case the probability p(drugs|black) is very slightly higher than the probability p(drugs|~black), so the police feel that they are justified in using ethnicity as a feature for predicting drug use (it "works"). However, the probability p(drugs|black) many still be very much lower than the probability p(~drugs|black) a probability that is never considered. As we see in the previous slide the rates are around p(drugs|black) = 0.09 vs. p(~drugs|black) = 0.81. If the police blindly stop only black people, they are disadvantaging over 80% of the people they stop.

This is racism in a nutshell. People are disadvantaged not because of their actions, but because of a feature they share with somebody else who perpetrated a crime. A fundamental property of our modern morality is that people should only be judged by their own actions, and should only be punished for their own actions. Under the banner of "it works" machine learning and data mining can be responsible for horrible acts of discrimination when only their performance for a given task is evaluated and not whether or not their actions are fair.

Assuming that tall people play basketball may work better than assuming that short people play basketball, but you'll still assume that a lot people play basketball who don't. If that assumption carries a negative consequences must be taken into account. We end up disadvantaging people purely on the basis of a future they share with the people whom it if fair to disadvantage.

what if we forbid racial profiling?

Disallow the use of gender, ethnicity, sexual orientation etc. as features in sensitive ML tasks

What about: postcode, hobbies, average salary, mode of transport, etc.

What about companies: how do we police Google, Facebook, Yahoo?

38

Can we solve the problem by simply disallowing people the use of these sensitive features in datamining applications? The problem is that many other features are highly correlated. If we combine postcode, income, hobbies and music taste (and perhaps the same values of people close by), we can end up with a perfect predictor for the sensitive values. And since the ML algorithm is optimised for what works (in a very narrow sense) rather than what's fair, we will still see algorithms that discriminate.

Moreover, while we can police some institutions, we cannot police foreign companies. If we limit only domestic companies and government institutions, we are just creating an advantage for the institutions we can't control.

What if

the data is a fair representation of the population

and

the predictions are accurate

and

we've correctly used Bayes rule?

39

actions versus predictions

It is fundamentally unfair to hold an individual responsible for the the actions of others that share their attributes.

Everybody has the *right* to to be judged on their own actions.

"hold responsible":

subject to a traffic stop, not give parole, search at an airport, not give a credit card, make it more difficult to get a job.

---

feedback loops

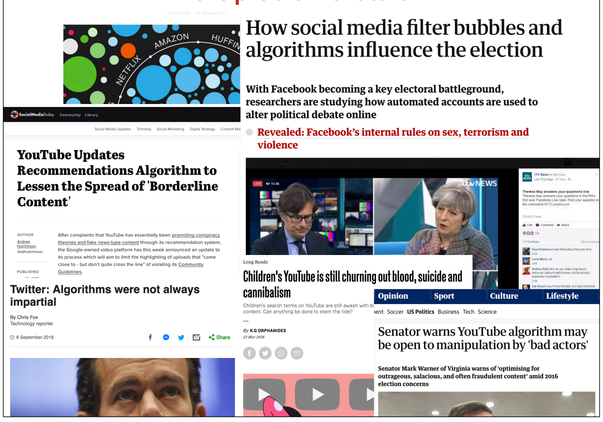
Offline learning doesn't stay offline.

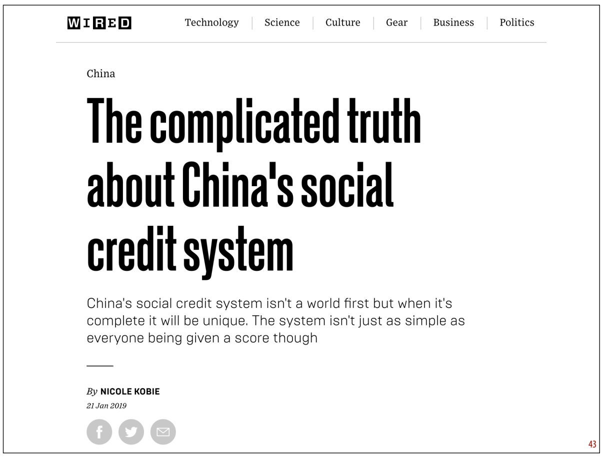
Predictions become actions, that reinforce existing biases from the data.

It's not just about whether the predictions are accurate. It's about whether the **actions are fair, and effective**.

---

the problem of scale





thank you for your attention

mlcourse@peterbloem.nl