COMP9414: Artificial Intelligence

Lecture 6c: Data Science and Ethics

Wayne Wobcke

e-mail:w.wobcke@unsw.edu.au

Overview

- Problems
  - Overfitting
  - Bias and Discrimination
- Methodology
  - Feature Engineering
  - Local Contextual Assumptions
  - Aggregating and Disaggregating Datasets
  - Validation

What Data Science is Not (A Caricature)

- Choose a complex concept/statistic/indicator to measure
  - Poverty/wealth indicators, food security map
- Choose a number of large-ish datasets
  - Mobile phone data, satellite data, admin data, survey data
- Choose a number of “covariates” in addition
  - Nighttime lights, land use, etc.
- Throw all data into standard method in R/Python, $\cdots$
  - Decision Trees, Random Forests, XGBoost, Neural Networks, $\cdots$
- Gives mixed results (to the extent validated $\cdots$)

Problem: Overfitting

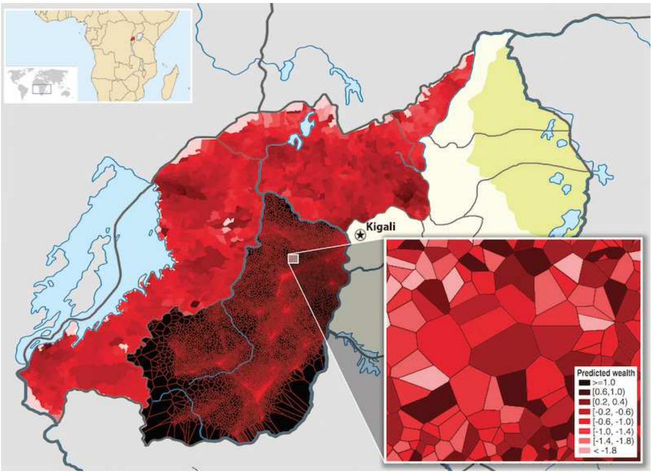
Overfitting = Fit given data too closely and not work in other contexts

Example: How not to measure wealth index (Blumenstock et al. 2015)

- Mobile phone data with 5088 features and 856 labelled examples
- Choose features based on whole dataset (not training set)
- Don’t consider what is Rwanda-specific about this data
- Use non-standard methodology drawn from another paper
- Ignore sensible (human-generated) baselines
- 5-fold cross-validation produces 5 models, not one

Claim(?): Many neural network/deep learning models overfit

Overfitting



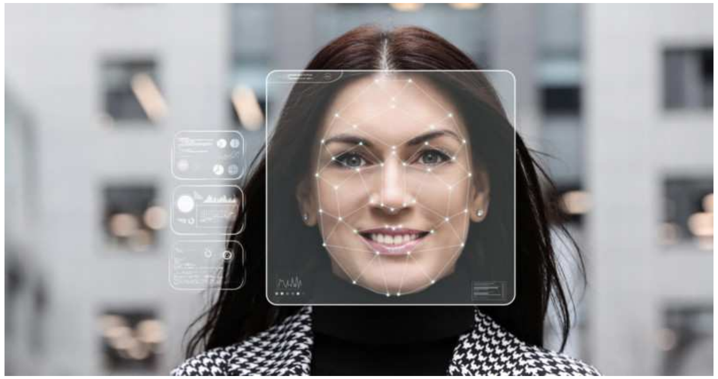
Problem: Bias and Discrimination

Bias = Propensity for method to generalize (good or bad)

- Dataset not representative of population
  - Only people in areas with phone towers have phones
  - Only people who are literate can send text messages
  - Only poorer people need “access” to phone credits
- Learner generalizes “wrong” features
  - White background (only pictures of snow leopards are in winter)
- Learner “misses” relevant features
  - Seasonal effects of population movement (food shortages)

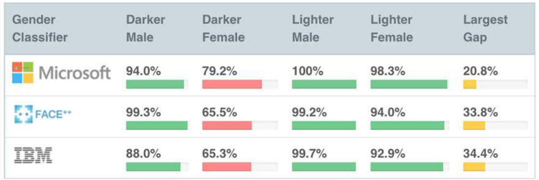
Bias (in machine learning) can lead to (unethical) discrimination

Clearview AI



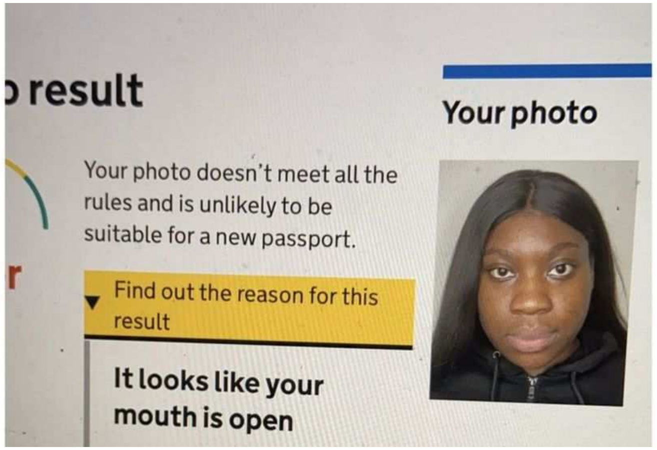
Facial Recognition Bias

Facial Recognition Bias



Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

UK Passports Discrimination



o result

Your photo

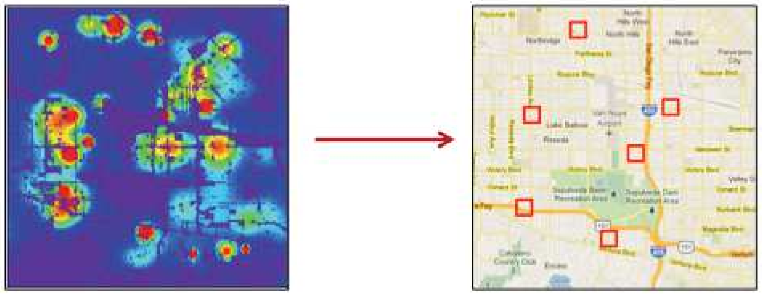
Your photo doesn't meet all the rules and is unlikely to be suitable for a new passport.

▼ Find out the reason for this result

It looks like your mouth is open

Wrongful Arrest Discrimination

Predictive Policing Discrimination



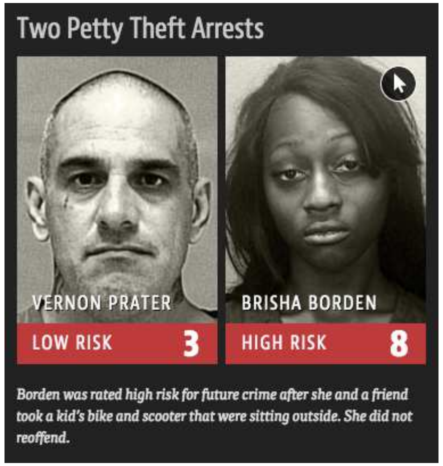
TACTICAL AMBIGUITY

rear-view mirror heat map

TACTICAL CLARITY

forward-looking PredPol boxes

Recidivism Rating Discrimination



Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

Data Science Methodology

- Methodology: In statistics/machine learning textbooks
  - Methods, models, theorems, estimators, techniques, tools
- Meta-methodology: Knowledge and practices that support this
  - How is it decided what "concepts" to measure?
  - How is it decided how these concepts are defined?
  - How is it decided how these concepts are measured (what data)?
  - How is robustness or reliability of results checked?
  - How are the results validated (internal and external)?
  - How do the results influence policy/decision making?

Lack of emphasis in textbooks, but very important to learn

Human Element of Data Science

Essential when data is limited in quality, quantity (most of the time)

- Human suggests relevant features
  - Protest less likely to be violent if venue private
  - AfPak ontology of events of interest to conflict progression
- Human defines useful indicators
  - Village is safe if market is open at night
- Human validates model output
  - Check agreement with model on 15% random sample
  - Verify main features used by the model
  - Define baseline for comparative performance
  - Cross check model output with other datasets

Feature Engineering

Example: Mobile Phone Data includes location of cell towers

- Location is Angkor Wat and time is 1 day ⇒ tourist?
- Or, journey "similar to" typical tourist trips ⇒ tourist
- Location is shopping centre ⇒ shopping (if not home)?
- Most frequent called person ⇒ spouse? (if married)
- Spouse ⇒ opposite gender (use as a check)
- Location is port and truck driver ⇒ shipment
- Destination(s) of truck ⇒ type of shipment?

Methodology: Emphasis on dealing with multiple levels of uncertainty

Local Contextual Assumptions

Food Consumption Score

- 2100 calories per day estimated by weighting food types
- Weights motivated but oil and sugar “need adjustment”
- Locally validated (seasonal effects, local variations)
  - North Sudan vs South Sudan
  - Seasonal variation in Cameroon
- Correlate with other measures (admin data, surveys)

Ideally measures capacity(?), not behaviour
Impossible to learn even with a lot of data, need expertise

Combining Datasets

Use of only one type of data is insufficient for many purposes

- Especially social media data (Twitter, Facebook)
- Especially with complex metrics and indicators
  - Population health using images of hospital carpark
  - Rainfall locations and amounts using satellite data
- Need triangulation/corroboration, not increased uncertainty
  - Need to “correlate” independent data sources

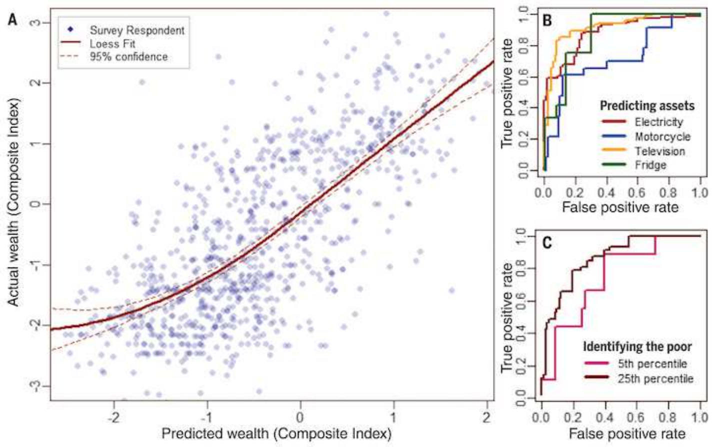
Pipelined Processes

- ADB poverty mapping (land use → regression)
- Errors in Phase 1 most likely systematic, not random
  - Gauss-Markov assumptions do not hold
  - Need to empirically estimate rather than use theory
  - Relies on “ground-truth” dataset
- Methods vs models
  - Works (better) for Philippines, not Thailand: why?
  - Tradeoff generality of method and “local validation”

Slicing and Dicing

- Data may only be reliable in certain contexts
  - May be able to determine event occurrence, not details
  - Sentiment analysis notoriously inaccurate
- May want to analyse subgroups by region, status, etc.
  - “Big data” can soon become “small data”
  - Need statistical methods to assess reliability
  - Map quality of data to quality of resulting decision

Validation



Conclusion

Is data fit for (what) purpose?

- No model is ever perfect (especially learned models)
- Statistical correlations are usually very weak
- Contextualize models to local circumstances
- Cross check model outputs with other datasets
- Express uncertainty associated with conclusions/decisions
- “Big data” methods can provide “early warning” signals
- Complement traditional measures with different time scales
- Continually validate models as assumptions vary