CMT307: Applied ML
Session 8

Ethics and bias
Recap

Jose Camacho Collados

Outline

- Guest lecture
  - Dr Federico Liberatore: “Applications of Data Science in Policing”
- Ethics and bias
- Coursework
- Recap
- Mid-module feedback
- Second semester projects

Ethics and Bias

Ethics and bias

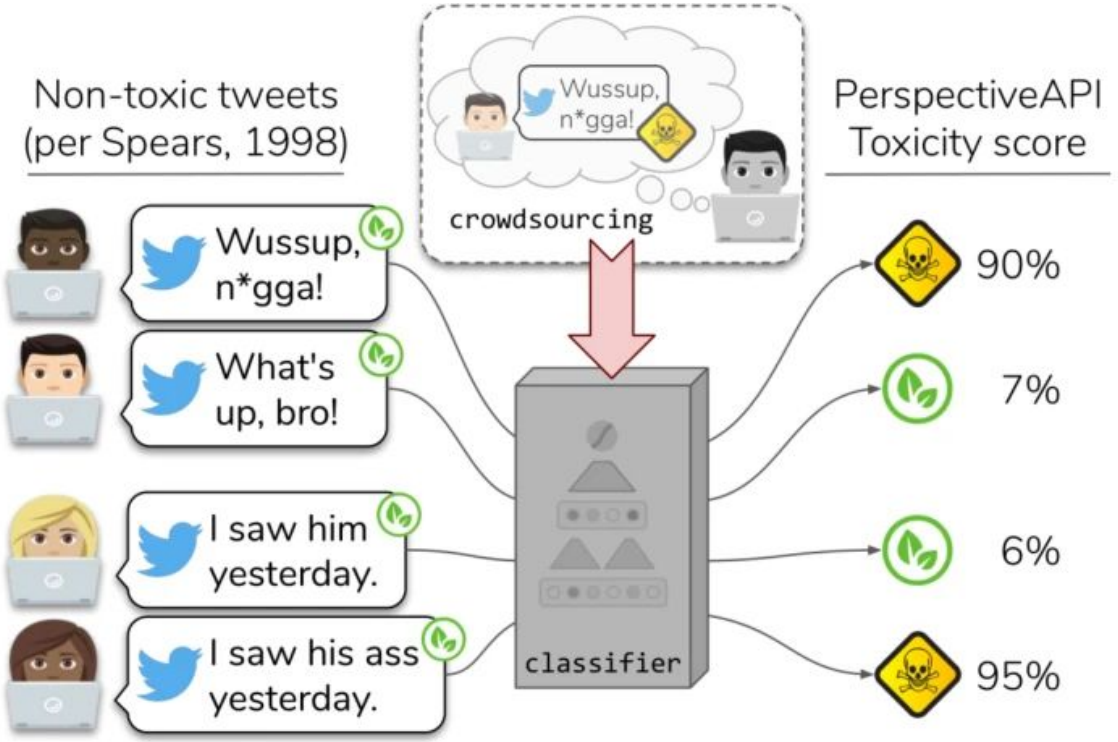
- It is important that we use Machine Learning **responsibly**, and that we are aware of its implications (and limitations).

- All machine learning algorithms contain some kind of **bias**.

- Deploying ML models (with bias) in the wild without mitigating their effect usually causes some kind of **discrimination**.

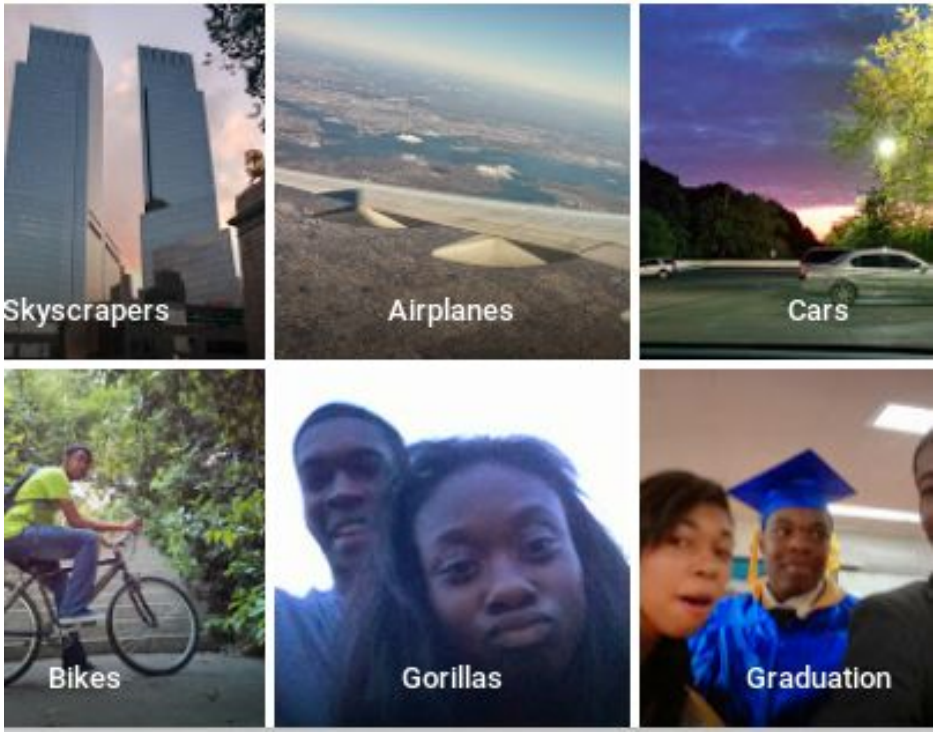
- Unfortunately, there is **no general recipe** to fight this problem, but the only way to find a solution is to first understand its limitations.

Bias in Machine Learning: Examples



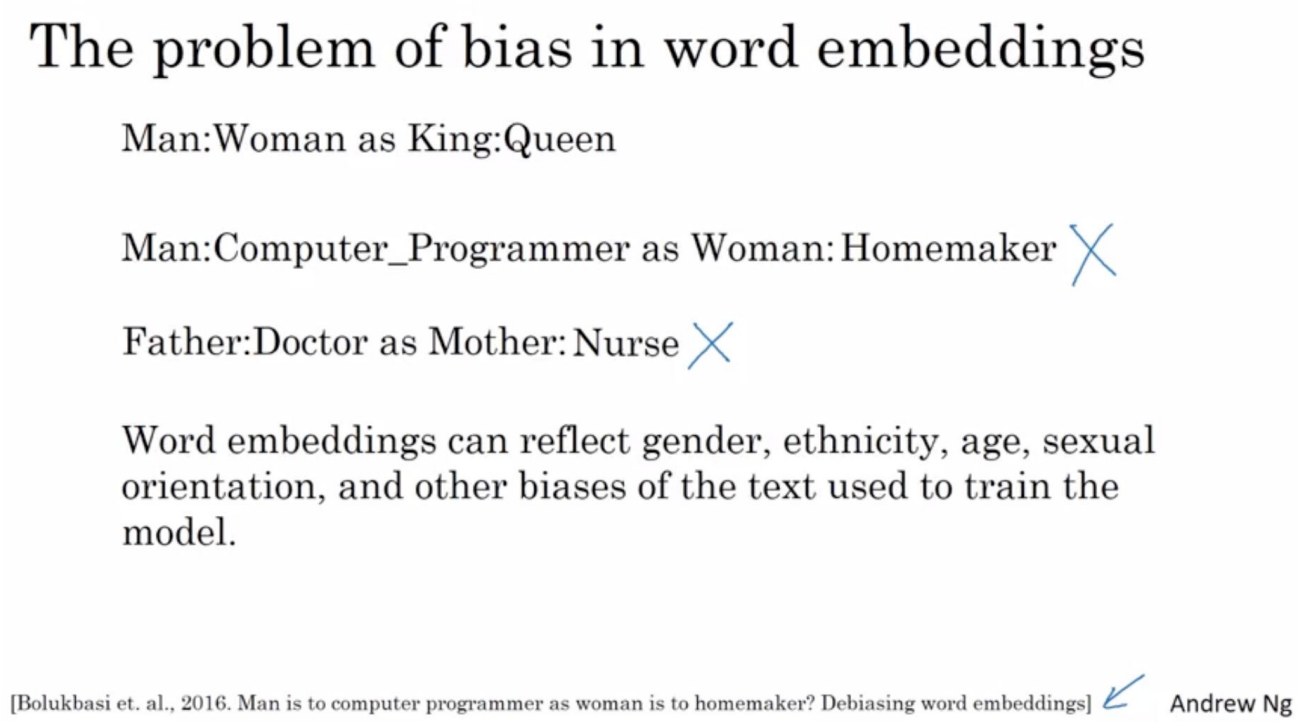
Slide credit: *Techcrunch*

Bias in Machine Learning: Examples



Slide credit: *Techcrunch*

Bias in Machine Learning: Examples



Slide credit: *Andrew Ng ; Reference: Bolukbasi et al. (2016)*

Fairness

- In many cases, the issue comes from the **data** itself. A careful collection/curation of the data is crucial.

- Problem: Data usually contain **human biases**.

- The selection of the ML algorithm is also important, and there are **ways to limit the effect of bias**.

More about fairness:
https://towardsdatascience.com/a-tutorial-on-fairness-in-machine-learning-3ff8ba1040cb

Ethics and Bias: Conclusion

- There are many moral and **ethical questions** around the use of Machine Learning in real-world applications.

- It is essential that we **understand** that biases will be introduced before deploying ML models in real-world applications.

- This way, we can try to find the best way to **mitigate their effects**.

Coursework

Coursework: General information

- **Available in Learning Central:** Since Monday, October 28th
- **Submission date and time:** Tuesday, January 14th at 9:30am

Most of the exercises are highly practical (Python, sklearn, etc.)

Extra credit possible (optional - a bit more challenging).

You are allowed to use functions/code given during the course.

Coursework: General Advice

Try to be **pragmatic in Part 1**. The exercises are similar to others done during the course (included in the Python notebooks).

Then, in **Part 2** you can think a bit more out-of-the-box. **Original ideas** can be considered positively even if they don't translate in good results (as long as they are clearly motivated/justified in the essay).

Part 1

Clarification:

No need to deliver solution to the lab exercises given throughout the course. Only answer to the exercises described in the Coursework sheet.

*“Part (1) consists of selected homework *similar to the one* handed in throughout the course”*

Part 1- Theory

Three questions

Maximum of 100 words for each answer

Please answer with your own words

Part 1- Practice: Question 1 (Evaluation metrics)

Just need to write how you get to the final result by applying the formulas of each metric (replacing them with the values from the table).

No Python code can be used for this exercise.

Part 1- Practice: Question 2 (Wine dataset)

Regression (not classification)

No need to include the full code, only what is asked (it can be copy-pasted, screenshot or anything that would be visible)

Part 1- Practice: Question 3 (Hateval dataset)

No need to include the code, just explain the process you followed in the short report.

You will **not be evaluated based on the results** (e.g. you can get full marks if your report is well written and the process makes sense, even if your results are not very high; and you can get low marks even if you get the best results).

Part 2: General comments

You can use any **Python external library**, even if it wasn't used during the course.

You can be as **original** as you feel.

Try to write some **justification/motivation** in your report of "non-standard" steps you may take (in some cases it can be just one sentence, justified by scores in the dev set, etc.)

Part 2

Clarification:

In this context, features can have one or more dimensions (e.g. length of the sentence would count as a feature).

“Students should include at least three different features to train their model, one of them should be based on some sort of word frequency”

Any σther questions about the coursework?

Coursework questions: COMSC Stack Overflow

https://stackoverflow.com/c/comsc

Please check the posts in Stack Overflow before asking new questions about the coursework. Then, post all your questions here (rather than email). Also, you can help your colleagues by answering questions!

Add the tag ***cmt307*** (and optionally ***machine-learning***) to your question.

Christmas break

There will be no answers to questions about the module/coursework during the Christmas break (~ 20 December - 9 January) - You can still post in Stack Overflow (questions may be answered by a student, or later by a TA/instructor).

Please ask your questions before the break (ideally today!).

Other options:

- On Monday (December 16) I will have **office hours** at the usual time (14:30-17:00)
- There will be a final session on **Friday, January 9 at 11am** (1 hour) for final clarifications about the coursework (optional). It will take place on **S/2.11**.

Recap

First semester so far

Session 1: Basic introduction to Machine Learning + Basic Python + Data preprocessing

Session 2: Continue data preprocessing + Feature selection

Session 3: Machine learning evaluation (cross-validation, evaluation measures, etc.)

Sessions 4-7 (Dr. Yuhua Li): Linear machine learning models (overview + theoretical concepts + mathematical foundations)

Session 8 (today): Guest lecture + Ethics and bias

Mid-module feedback

Mid-module feedback

Please answer the following form to provide feedback about the module:

https://bit.ly/2E7o1j0

Any constructive feedback will be highly appreciated!

Module code: CMT307

Second semester projects

Second semester projects (I)

In the second semester there will be **group projects** (of around 5-6 students each).

We will **start at the beginning of the semester** (more information will come in due term) and end by the end of April (approx.)

Second semester projects (II)

There will be a **variety of projects** to select from.

In the meantime, if you have a concrete **idea** for a particular machine learning project to include, please send an email to me and Dr. Yukun Lai (laiy4@cardiff.ac.uk) with a brief description of the idea. Then, we can discuss its feasibility.

Hands on!

The rest of the time you can work on any of the Python notebooks from class (any of the sessions) or on the coursework.

We are available for questions during this time