



Applied Machine Learning & Big Data

Course information 2019

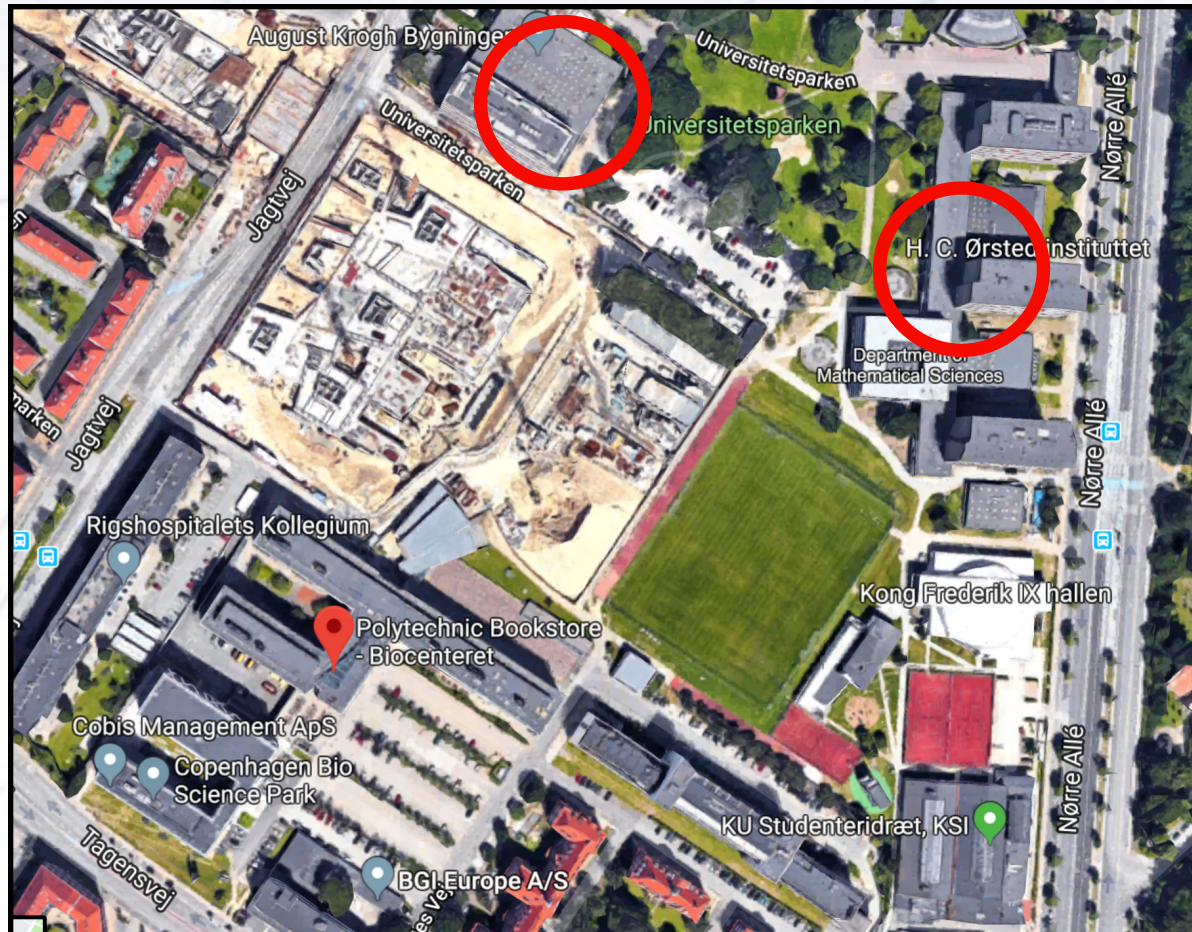
Troels C. Petersen, Joachim Mathiesen,
Brian Vinter & Adriano Agnello (NBI)

Lectures: August Krogh Building (Aud 1/3)

Monday: 13-14, Wednesday 9-10 (13-14)

Exercises: H. C. Ørsted Building (A110)

...rest of the time!



Additional locations

Troels' office
(building M, top floor)

Joachim's office
(building K, first floor)

Brian's office
(building C, top floor)

Blegdamsvej

Computers and software

We will program (mainly) in Python, but the choice is in fact your own. We will only provide data and inspiration.

We suggest that you use Jupyter Notebook, possibly on ERDA, but make sure that you can also run everything on your own laptop, so that you are not dependent on internet and ERDA access. We also recommend that you use GitHub/GitLab

Data files will typically be provided in CSV and/or HDF5 format, but others might be used.

We will be using many additional packages, introduced along the way, and surely you have your own favourites. Use them happily.

We hope to have enough power for everyone, but please try to have your computers charged, when coming to class, and/or bring an extension cord.

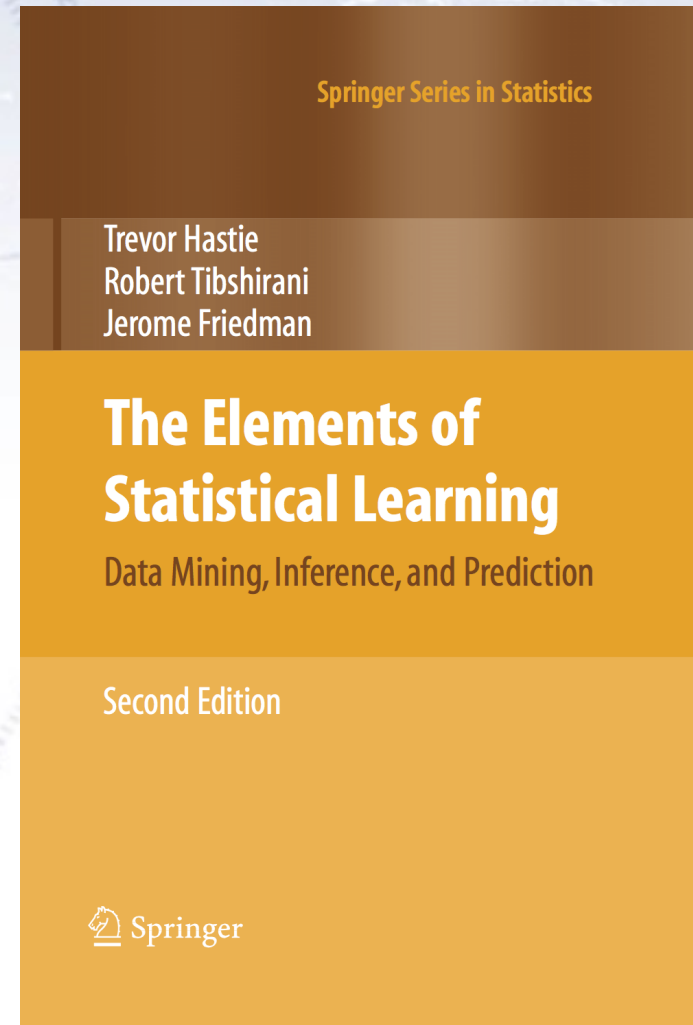
Literature

The main literature will be slides, notes, and references/links! However, we also wanted you to have a few more “solid” places to read comprehensively about ML.

“The Elements of Statistical Learning” (TEOSL) is a good read (though at times rather mathematical), and especially chapter 2 is a good introduction.

Part I, Chapter 5 of “Deep Learning” is also a great overview of ML and its ingredients.

Finally, Christopher M. Bishop’s “Pattern Recognition and Machine Learning” is good, but it is not available on the web (for free).



Expectations

We want (read: insist) this course to be useful to all of you! Therefore, please give us feedback (the earlier the better), if you have anything to add/suggest/criticise/alter.

However, it is also a VERY independent course in the sense that it is up to YOU, how much you get out of it. Consider it rather a project than a course!

The aim is to make you knowledgeable about the basics of Machine Learning, and being able to apply it to (suitable) data.

Problems?

If you experience problems in relation to the course, whatever their origin and nature, then write us!

We may not be able to do anything about it, but if we don't know about your problems, then I most certainly can not do anything about them.

We consider ourselves fairly large, as long as I feel that this largeness is met by sincerity and will.

But... you need to write us in the first place! That is your responsibility.

Projects / Exam

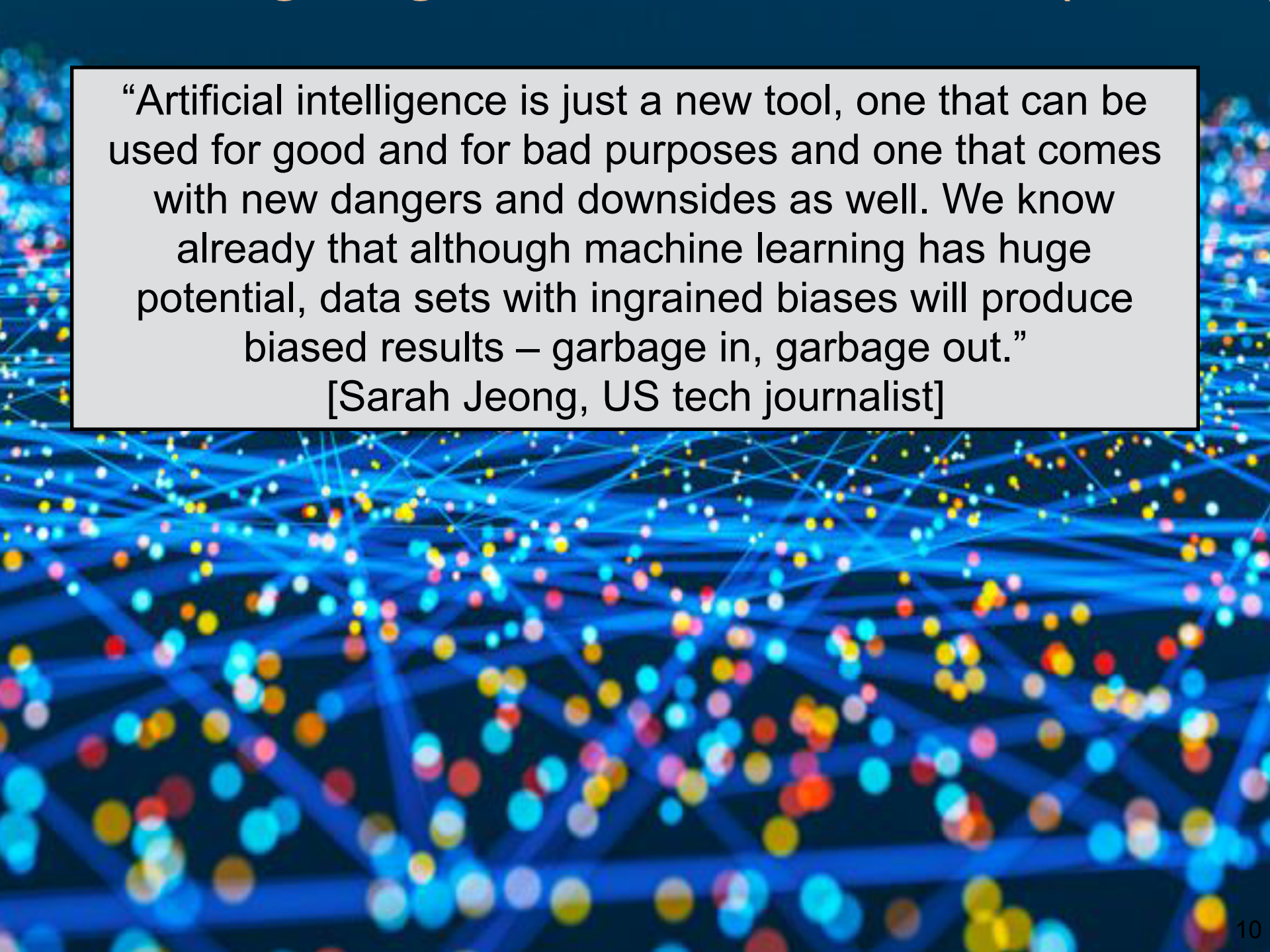
This course is in some sense a project!

We will be doing two projects:

- An initial small project on (common) data (1 week - 20% of your grade).
- A final larger project on data of your choice (3 weeks - 80% of your grade).

The initial project will be the basic application of ML to a simple data set, and we will evaluate your performance by applying your algorithm to a (to you unknown) validation data set.

The final project will be your main task, and can be the application of ML on anything that you like. You will all be presenting your results to each other, so that also others may learn from what you did (and didn't).



“Artificial intelligence is just a new tool, one that can be used for good and for bad purposes and one that comes with new dangers and downsides as well. We know already that although machine learning has huge potential, data sets with ingrained biases will produce biased results – garbage in, garbage out.”
[Sarah Jeong, US tech journalist]