Machine Learning applied to Planetary Sciences

PTYS 595B/495B

Leon Palafox

https://leonpalafox.github.io/MLClass/

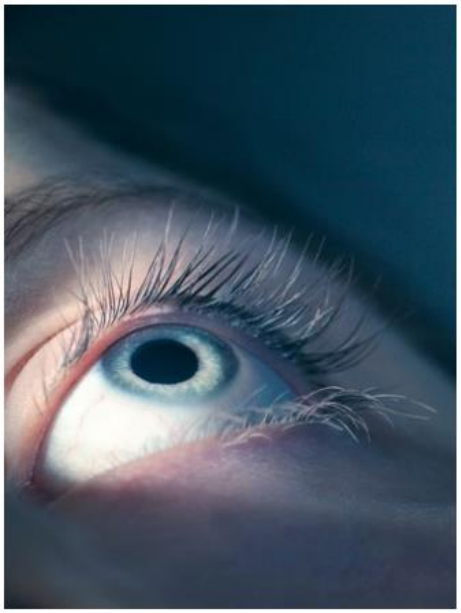
PSA

- Homework questions?
- Final Quiz:
  - Unsupervised Learning
  - Date: 12/7 (Last Session)
- Final Project
  - Due date: 12/14
  - Questions?

Today in the news

- TensorFlow is available in Windows.
- Deep Learning being used to prevent blindness in diabetics.

GOOGLE'S AI READS RETINAS TO PREVENT BLINDNESS IN DIABETICS



GETTY IMAGES

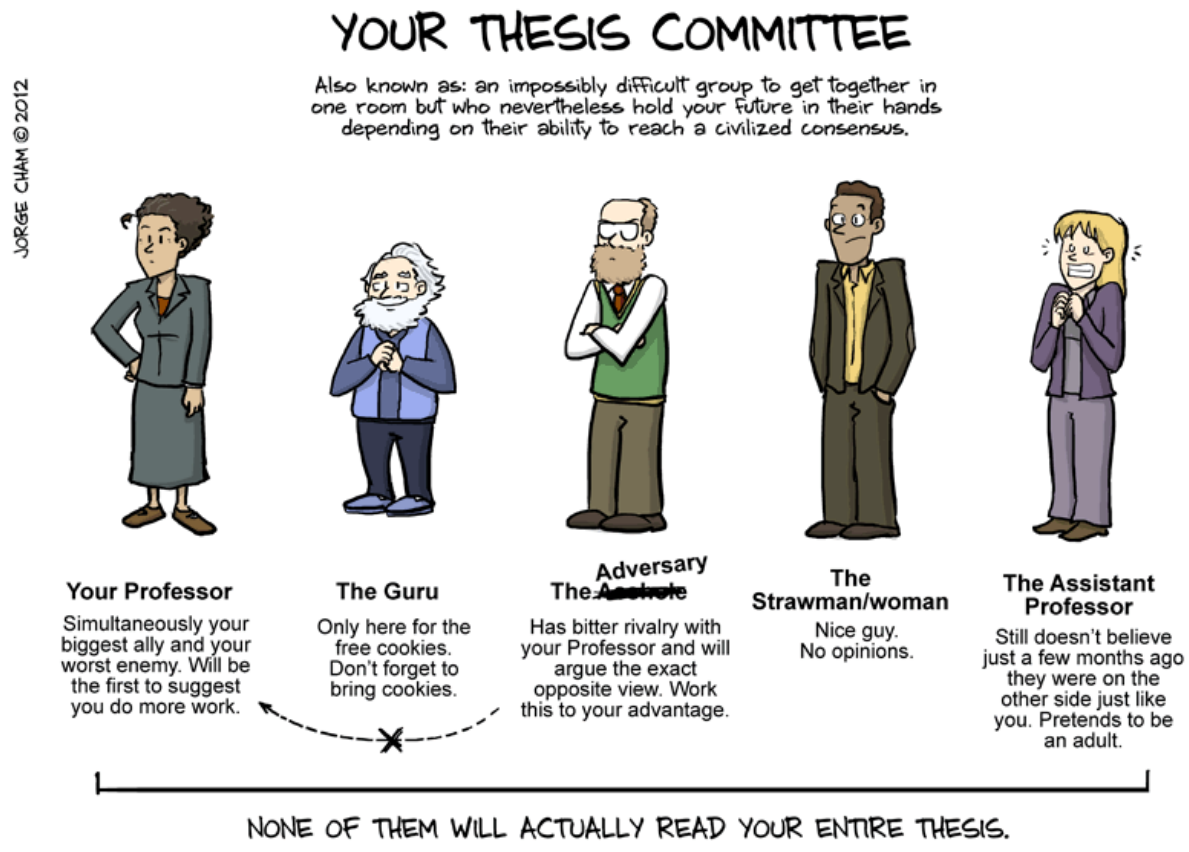
GOOGLE'S ARTIFICIAL INTELLIGENCE can play the ancient game of Go better than any human. It can identify faces, recognize spoken words, and pull answers to your questions from the web. But the promise is that this same kind of technology will soon handle far more serious work than playing games and feeding smartphone apps. One day, it could help care for the human body.

Feature decomposition

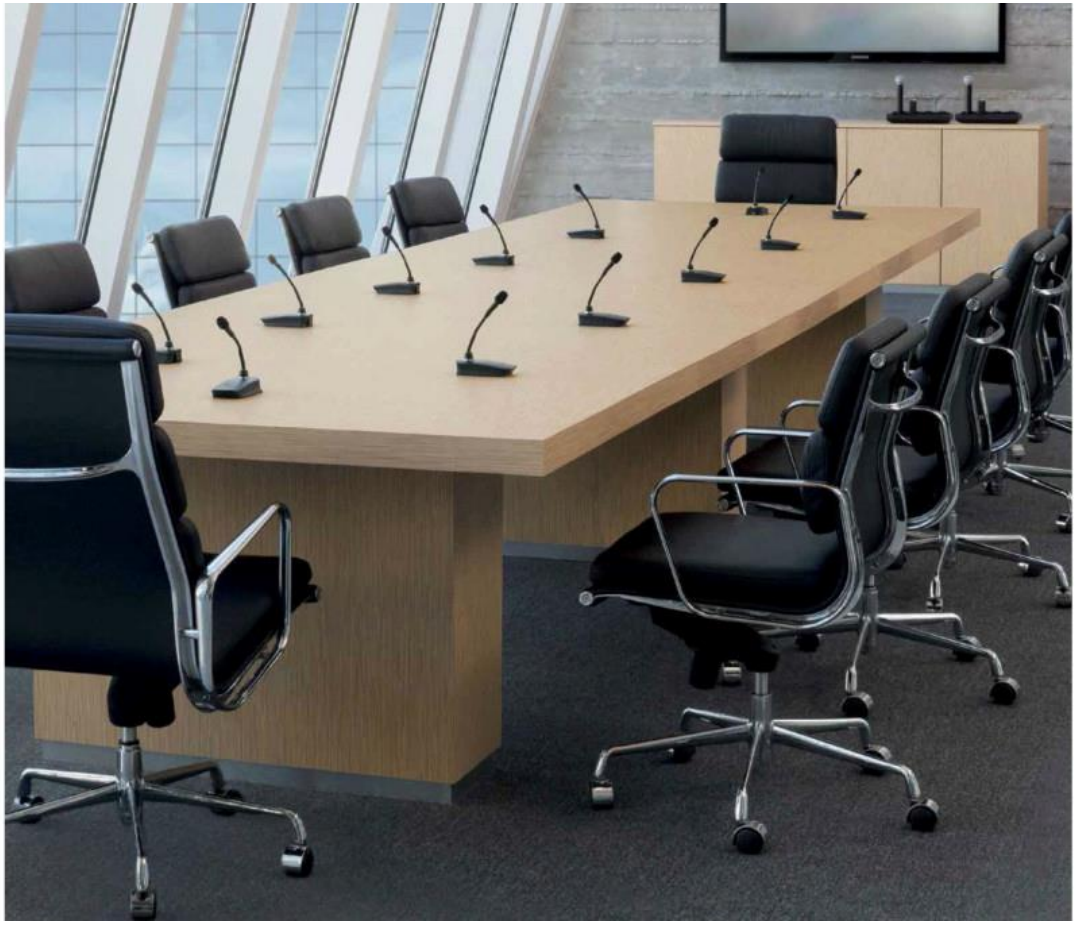
- When we have a very large number of features:
  - Many of them are redundant.
    - Speed in different dimensions
    - Different electrodes in an EEG signal
- People usually calculate the Correlation matrix.
  - It shows which features change in tandem.

~~Cocktail Party Problem~~

Thesis committee feedback problem



Five different signals



Formal definition of the problem

- Given s (sources), a mixing matrix A, we have the observed signal x

$$x = As,$$

- The un-mixing problem consists on trying to find both A and s

Some constraints

- A is impossible to recover if we have no prior information of the data.
- Given a permutation matrix P

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \quad P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

- If z is a vector, then Pz is permuted version of z.

Some constraints

- PW and W are impossible to distinguish. But this is of little importance.
- There is no way to recover the correct scaling.
- The data can’t be Gaussian (luckily enough, most data isn’t)

Computational Constraints

- ICA is way more intensive than PCA, since it does an optimization to find the right A and s.

- In general we MUST use whitening as well in ICA, remember that ICA has troubles with scales.

Matlab/Python

- sklearn.decomposition.FastICA
  - N_components = # of components
  - whiten = whether we use whitening or not (True/False)
  - W_init = Mixing matrix for initialization.
- http://research.ics.aalto.fi/ica/fastica/
  - There is no official Matlab package, but the author of the algorithm has made available the library.