Session 1

Basic introduction to Machine Learning and Python

Outline

- Module information
- What is Machine Learning
- Practical information
- Hands on!

Module Information

Module instructors

Yukun Lai

Jose Camacho Collados

Yuhua Li

Stefano Zappala (TA)

Steven Arthur (TA)

Module organisation

First semester

Basic Machine Learning concepts

+

Linear models

Second semester

Neural networks

Module plan: first semester

Session 1 (week 2, 10 Oct): Basic introduction to Machine Learning + Basic Python + Data preprocessing (Jose)

Session 2 (week 3, 17 Oct): Continue data preprocessing + Feature selection (Jose)

Session 3 (week 4, 24 Oct): Machine learning evaluation (cross-validation, evaluation measures, statistical testing, etc.) (Jose)

Session 4 (week 6, 7 Nov): Linear machine learning, support vector machines, kernels (Yuhua)

Session 5 (week 7, 14 Nov): Continue linear machine learning + Bias-variance tradeoff, no free lunch theorem (Yuhua)

Session 6 (week 9, 28 Nov): Decision trees, random forests, logistic regression (Yuhua)

Session 7 (week 10, 5 Dec): Meta-learning (boosting, bagging, stacking) + Application (Yuhua)

Session 8 (week 11, 12 Dec): Ethics and bias + Applications + Recap (Jose)

Courseworks

First semester

[Individual]

Design and implement a machine learning method for solving a problem.

Write a report.

Second semester

[Group work]

Design and implement a neural network for solving a problem.

Write a group report, give a presentation. Write an individual report.

What you are going to learn (and not) in this module

- Basic understanding of (supervised) Machine Learning
- A bit of theory
- (Mostly) Practical tips and hands-on experience

- Theoretical principles
- Mathematical formulation behind Machine Learning models

Class: time distribution

14:10-15:20 -> Lecture

15:20-15:40 -> Break

15:40-17:00 -> Exercises/hands-on

This time distribution may vary from class to class.

Lectures could include practical activities as well.

What is Machine Learning

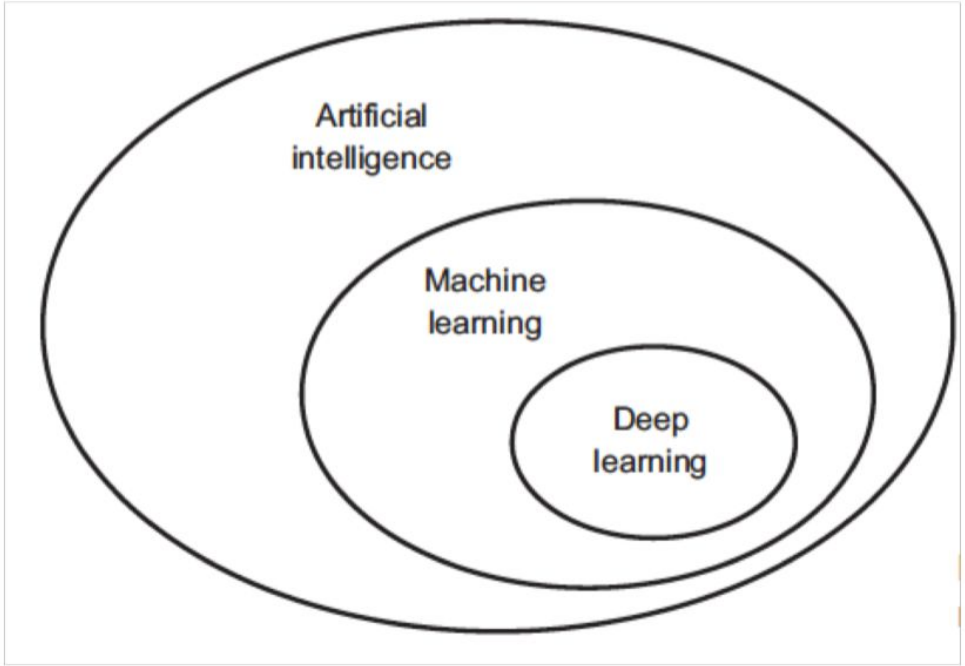
Machine Learning

Machine learning is…

*“is the scientific study of **algorithms and statistical models** that computer systems use to perform a specific task **without using explicit instructions**, relying on patterns and inference instead”* (Wikipedia)

*“Machine learning is an application of artificial intelligence (**AI**) that provides systems the ability to **automatically learn and improve from experience** without being explicitly programmed”* (Expert system, March 2017)

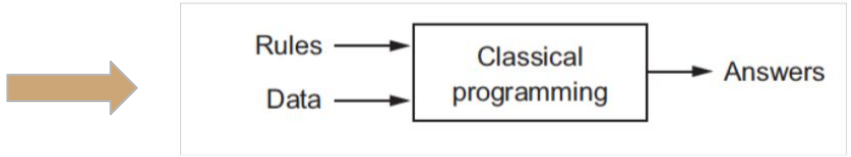
AI, Machine Learning and Deep Learning



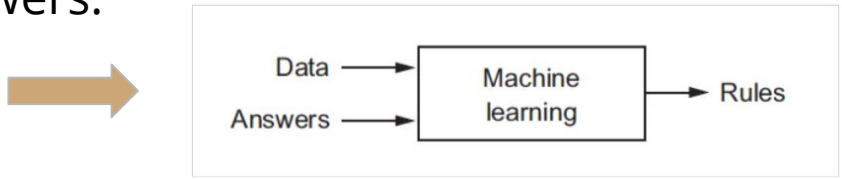
Slide based on *Deep Learning with Python*

Machine Learning vs. Symbolic AI

- **Symbolic AI:** humans input rules (a program) and data to be processed according to these rules, and out come answers.



- **Machine learning:** input data as well as the answers expected from the data and learn the rules automatically. These rules can then be applied to new data to produce original answers.



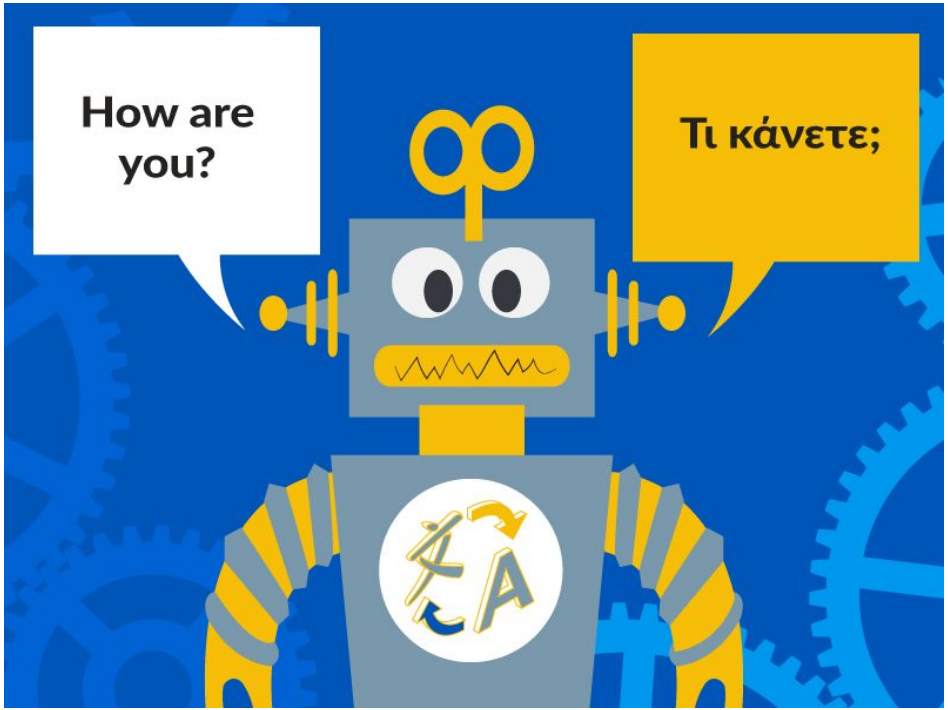
Machine Learning

A machine learning system is **trained** rather than explicitly programmed.

It is presented with (many) **examples relevant to a task**.

It finds statistical structure in these examples that eventually allows the system to come up with rules for automating the task.

Applications of Machine Learning



Machine Translation

Image credit: *Ciklopea*

Applications of Machine Learning

Games

(chess)

Image credit: *Reuters*

Applications of Machine Learning

Question answering

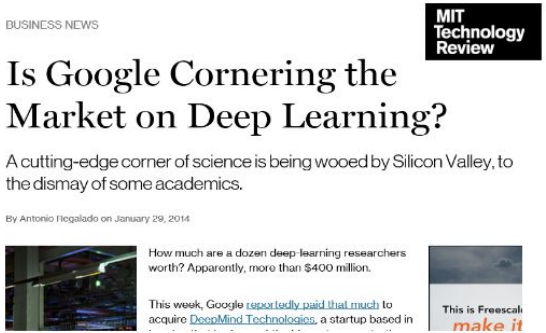
Image credit: *Jeopardy Productions*

Applications of Machine Learning

Self-driving cars

Vimeo screencap

Deep learning


BUSINESS NEWS

MIT Technology Review

Is Google Cornering the Market on Deep Learning?

A cutting-edge corner of science is being wooed by Silicon Valley, to the dismay of some academics.

By Antonio Regalado on January 29, 2014

How much are a dozen deep-learning researchers worth? Apparently, more than $400 million.

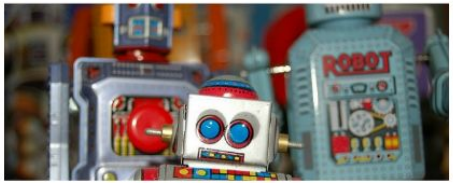
This week, Google reportedly paid that much to acquire DeepMind Technologies, a startup based in

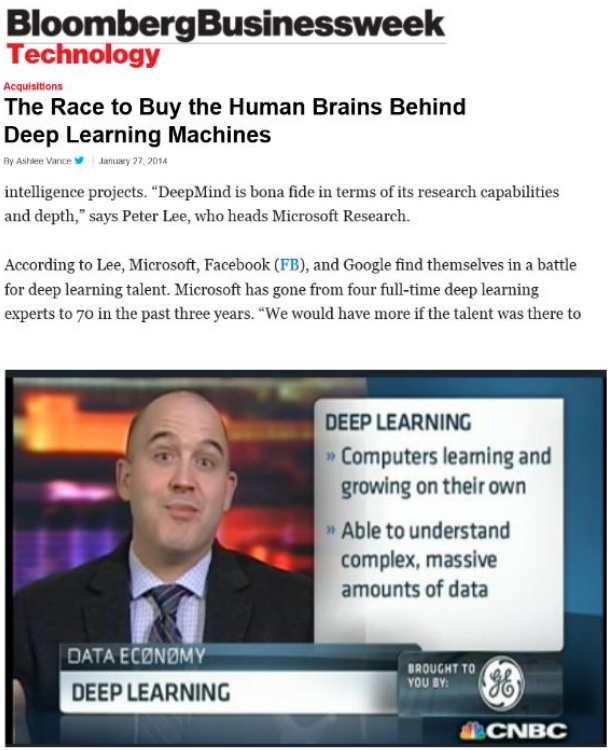

WIRED GEAR SCIENCE ENTERTAINMENT BUSINESS SECURITY

INNOVATION INSIGHTS | community content | featured

Deep Learning's Role in the Age of Robots

BY JULIAN GREEN, JETPAC 05.02.14 2:56 PM




BloombergBusinessweek Technology

Acquisitions

The Race to Buy the Human Brains Behind Deep Learning Machines

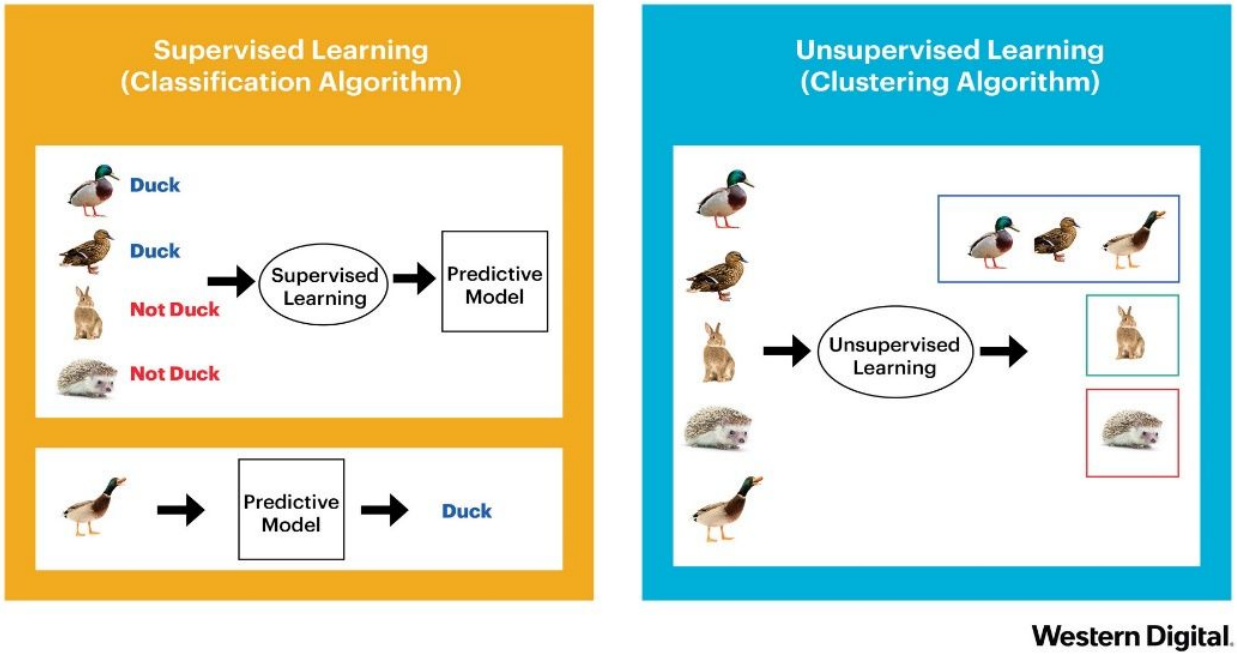
By Ashlee Vance | January 27, 2014

intelligence projects. "DeepMind is bona fide in terms of its research capabilities and depth," says Peter Lee, who heads Microsoft Research.

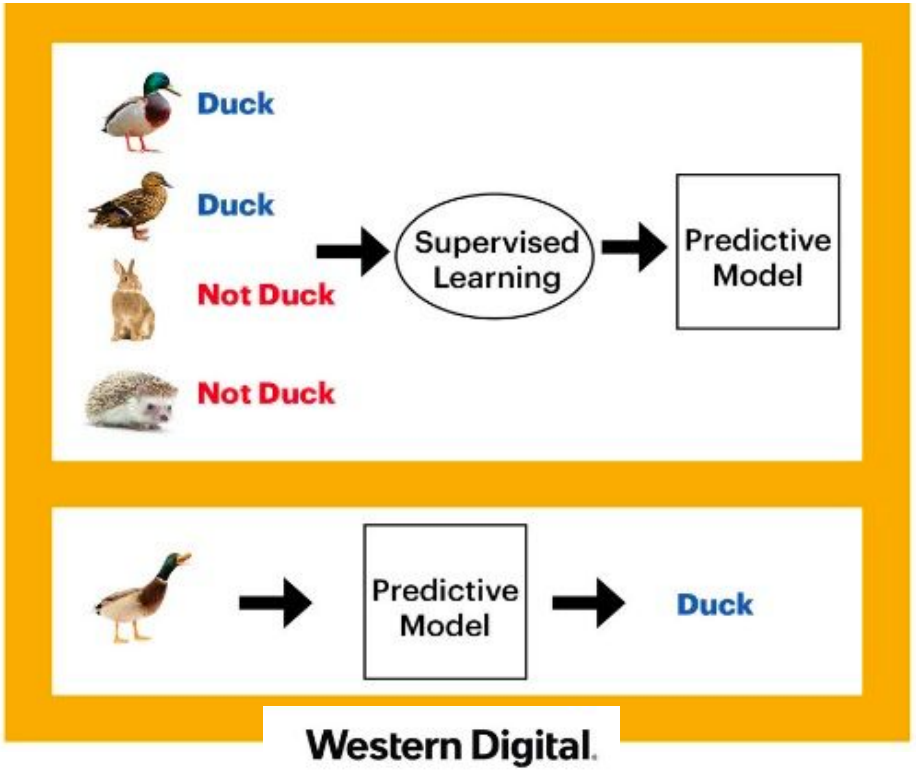
According to Lee, Microsoft, Facebook (FB), and Google find themselves in a battle for deep learning talent. Microsoft has gone from four full-time deep learning experts to 70 in the past three years. "We would have more if the talent was there to

Slide credit: *Eric Eaton*

Machine Learning: Supervised vs. Unsupervised

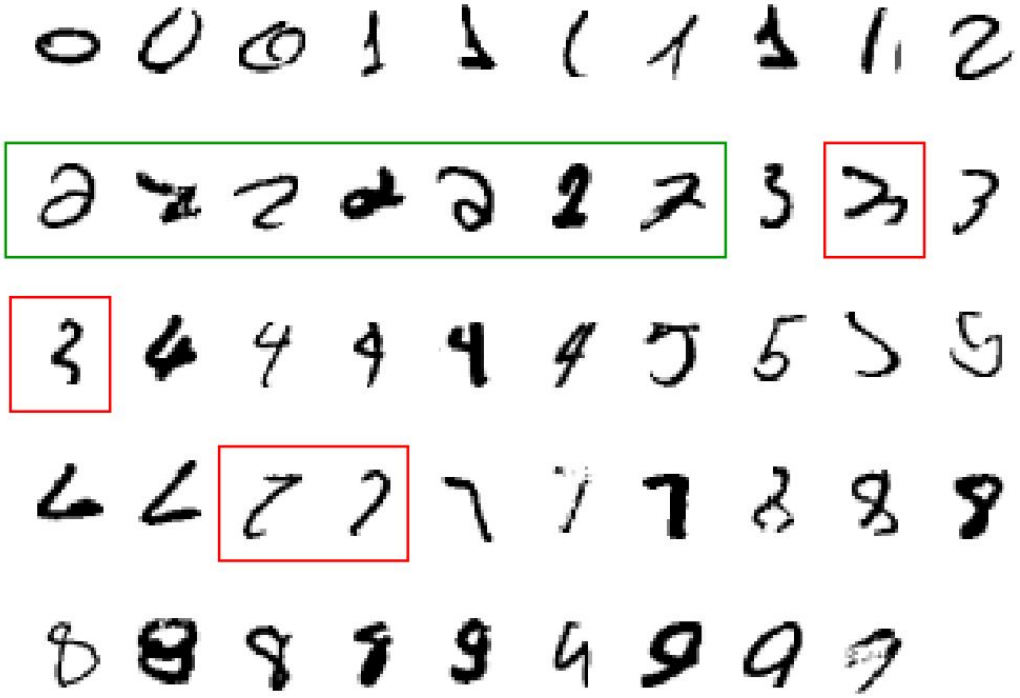


Supervised Learning: Classification



Main focus of this module!

Example: Computer vision

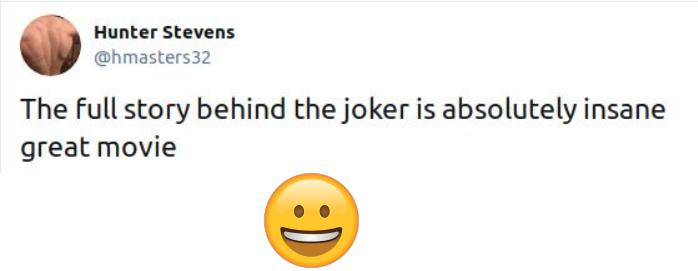


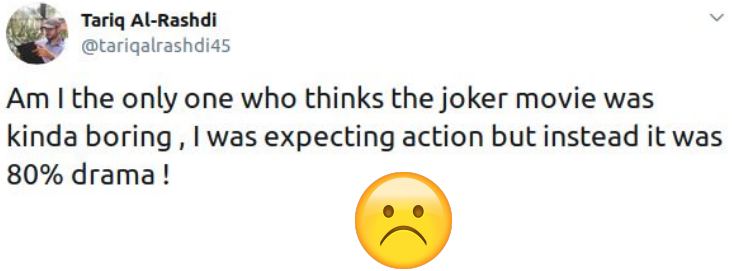
Slide credit: *Eric Eaton/Geoffrey Hinton*

Example: Sentiment analysis

We want to understand whether the audience has liked or not a new movie.

There are thousands of opinions from users in social media, fora, etc.





We can develop a **machine learning classifier** to understand users' opinions automatically!

Example: Sentiment analysis

- I liked the movie 
- The movie was awesome 
- It was quite boring 
- I enjoyed the movie 
- It was great! 
- The main actor was terrible 

...

Example: Sentiment analysis

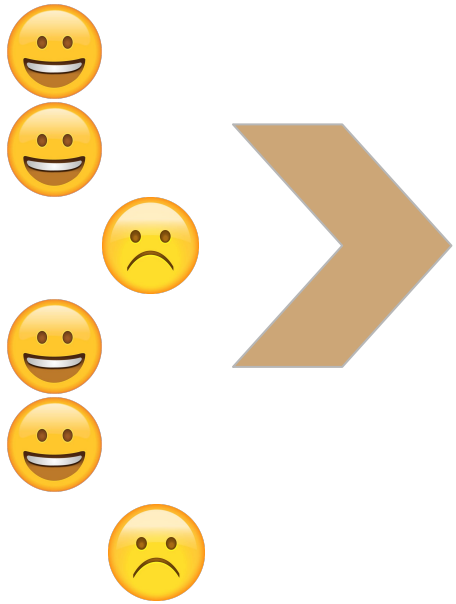
- I **liked** the movie 
- The movie was **awesome** 
- It was quite **boring** 
- I **enjoyed** the movie 
- It was **great**! 
- The main actor was **terrible** 

...

Example: Sentiment analysis

- I **liked** the movie
- The movie was **awesome**
- It was quite **boring**
- I **enjoyed** the movie
- It was **great**!
- The main actor was **terrible**

...



Train a machine learning classifier

Example: Sentiment analysis

The movie was great?  **or** **?**

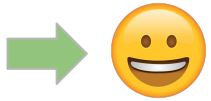
I got bored  **or** **?**

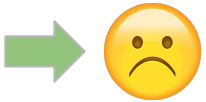
Example: Sentiment analysis

The movie was **great**?  **or** **?**

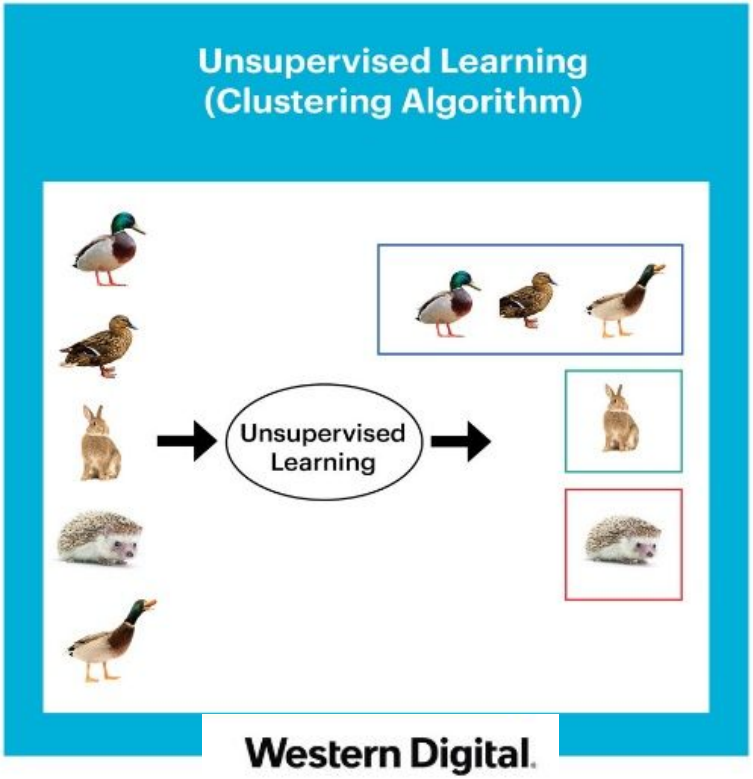
I got **bored**  **or** **?**

Example: Sentiment analysis

The movie was **great**? 

I got **bored** 

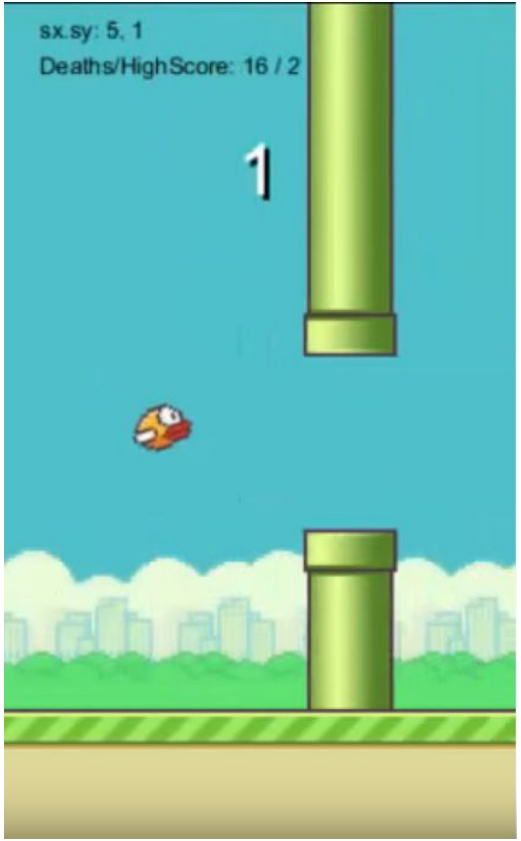
Unsupervised Learning: Clustering



Reinforcement learning

Usually considered as a different type of learning.

Rewards from sequence of actions



https://www.youtube.com/watch?v=rvy1ZS2RwBs

Practical Information

Python libraries

Essentials for this module:

- **Numpy** -> Mathematical functions, vectors
- **Sklearn** (scikit-learn) -> Machine Learning
- **Keras (Tensorflow)** -> High-level API for deep learning

Very useful:

- **NLTK** -> Statistical text processing
- **Spacy** -> Natural language processing
- **Gensim** -> Vectors, topic modeling
- **Pytorch** -> Deep learning library
- **Pandas** -> Data analysis
- **Matplotlib** -> Visualization

Google Colab 

Colaboratory is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud.

It contains all main Python libraries required for this course.

Suggested for this module!

Google Colab 

https://colab.research.google.com

It follows the same block-based format as **Jupyter Notebooks**. Most assignments will be delivered with this format.

Google Drive is used as storage unit.

It has free **GPU**-time which can be very useful for the second semester.

Colab: Some useful introductions/tutorials

- Introduction: https://colab.research.google.com/notebooks/welcome.ipynb#
- Google Colab tutorial - An introduction to Google Colab, How to upload files? How to connect to Google Drive and resume training.
- Gettings started (blog post): https://towardsdatascience.com/getting-started-with-google-colab-f2fff97f594c

Offline alternatives to Google Colab

Install all packages + **Jupyter Notebook** in your laptop/computer.

If no administrator (sudo) rights in computer -> **Atom** (pre-installed in some MSc computers) and virtual environments.

The School advices the use of **Pipenv** as virtual environment: https://github.com/pypa/pipenv

Bibliography (books)

Machine Learning: The Art and Science of Algorithms that Make Sense of Data, by Peter Flach.

Understanding Machine Learning: From Theory to Algorithms, by By Shai Shalev-Shwartz and Shai Ben-David: http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/

Hands-On Machine Learning with Scikit-Learn and TensorFlow, by Aurélien Géron

Deep Learning with Python, by Francois Chollet (2nd Semester)

References available in learning central

Online resources/courses

- Machine learning from University Wisconsin-Madison (includes lecture notes): https://github.com/rasbt/stat479-machine-learning-fs19/blob/master/README.md
- Coursera: https://www.coursera.org/learn/machine-learning
- Machine learning classifiers: https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623
- List with more resources! https://medium.com/ai2-blog/how-to-get-up-to-speed-on-machine-learning-and-ai-a0fd923d4169

School's private Stack Overflow

https://stackoverflow.com/c/comsc

Add the tags ***cmt307*** and ***machine-learning*** (if related to machine learning) to your question.

What to do now?

Refresh topics:

- Python programming
- Basic mathematics (linear algebra, probability, etc.)
- Linux (using the terminal, etc.)

Getting familiar with Colab (**Jupyter notebooks**, etc.) and Google Drive. Best way to do this: going over the exercises!

Virtual environments (**Pipenv**) - essential if not using Colab and no sudo

(Short) online tutorials to refresh topics

http://tiny.cc/q5vsdz

Compilation of online tutorials for **programming** (Python), basic **mathematics** (linear algebra and probability/statistics) and the usage of **terminal and virtual environments**.

Hands on!

Python notebook with exercises
about **data preprocessing** and **vector manipulation** available at Learning Central