CMT307: Applied ML Session 14

Recap + guest talks

Jose Camacho Collados

Outline

- **Update course** (coursework + group projects)
- **Recap**
- **Guest talks:** Career opportunities

Update course - coronavirus

Update course - coronavirus

Please check the **announcement in Learning Central** for all the details.

Main points:

- All deadlines extended one week.
- Presentations to be pre-recorded.
- Online group meetings among members and supervisor.

Office hours next Wednesday (March 25)

- Usual time: 2:30pm to 5pm
- Pre-arranged time via email
- Online meetings in Teams (can be more than one student at a time)

Recap

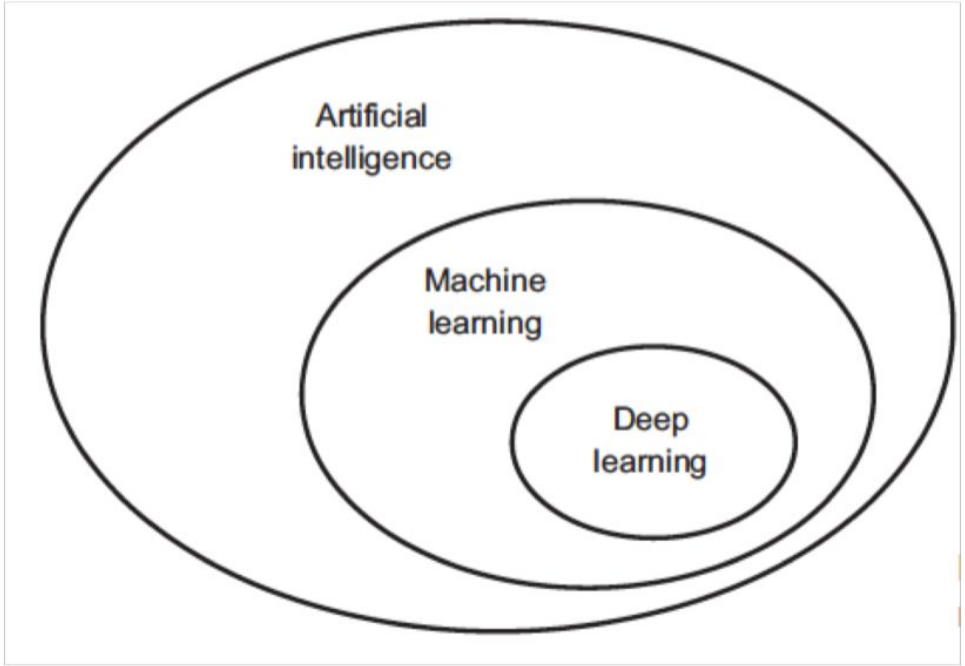
First semester: content

Session 1-3: Introduction to Machine Learning + Data preprocessing + feature engineering/selection + evaluation

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

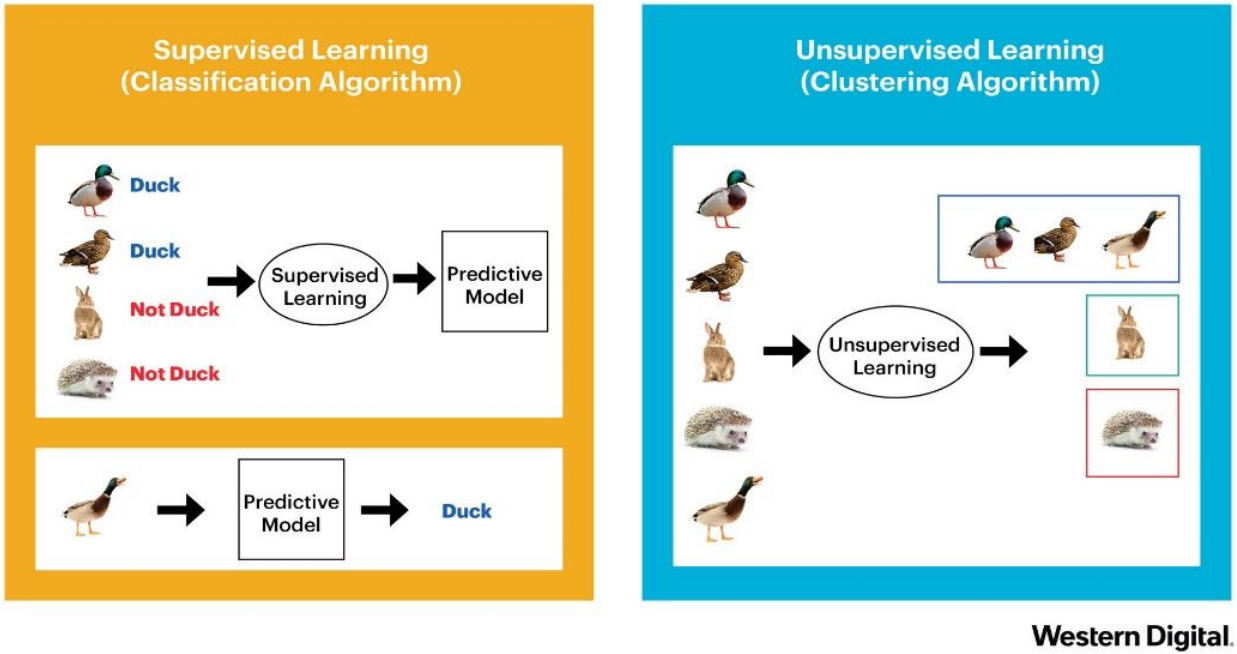
Session 8: Ethics and bias

AI, Machine Learning and Deep Learning

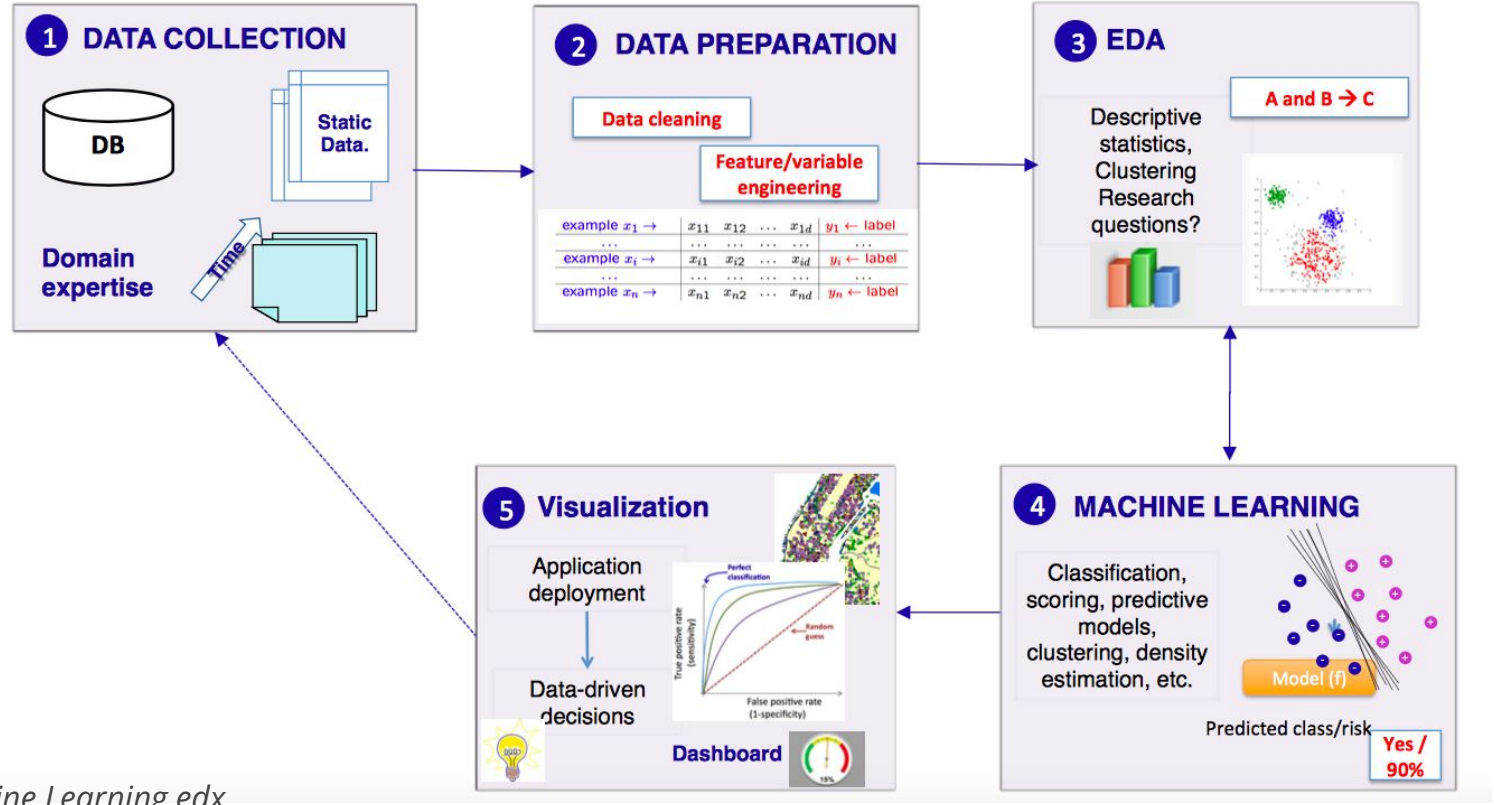


Slide based on *Deep Learning with Python*

Machine Learning: Supervised vs. Unsupervised

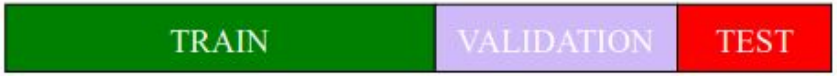


Machine Learning pipeline



Slide credit: *Machine Learning edx*

Train, validation and test



Datasets are usually split into three parts:

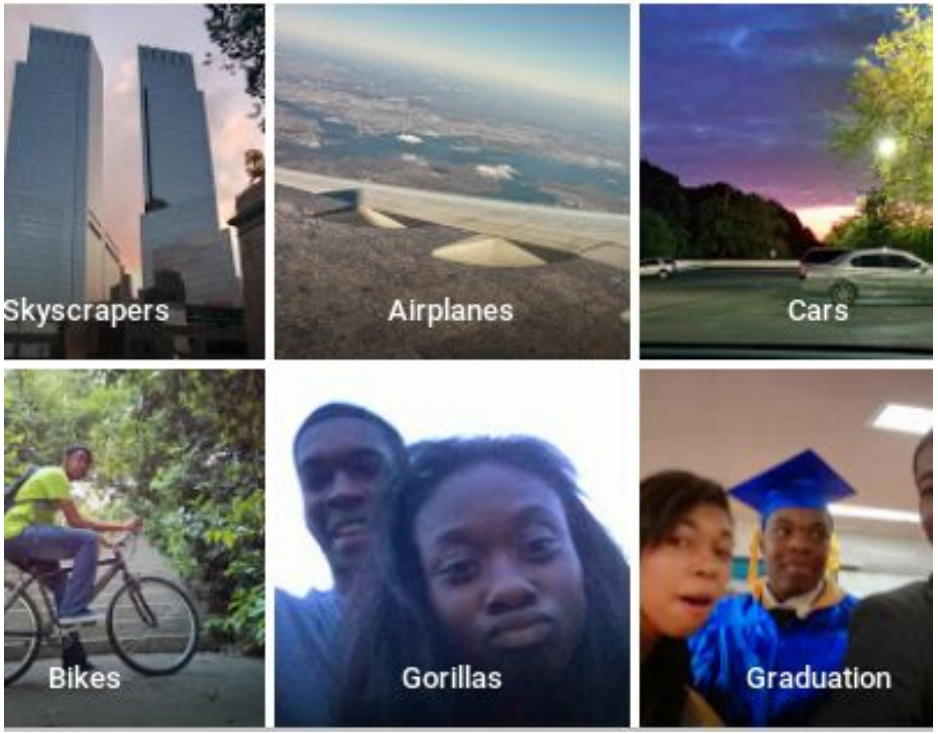
- **Training set:** Set of instances (examples) used to train a model.
- **Development (validation) set:** Instances that are used to train the parameters of your model -> Useful to avoid overfitting!
- **Test set:** Split where the model is evaluated (only used once!).

Slide credit: *Machine Learning edx*

Supervised Learning

- Regression
  - Linear regression, polynomial regression
  - Ridge regression, LASSO, elastic net
  - Fuzzy logic systems (FL)
  - Artificial neural networks (ANN) / deep learning
  - …
- Classification
  - Logistic regression
  - Support vector machines (SVM)
  - Bayesian network
  - K nearest neighbours (K-NN)
  - Decision tree (DT)
  - Ensemble learning
  - …

Bias in Machine Learning: Examples



Slide credit: *Techcrunch*

Second semester so far

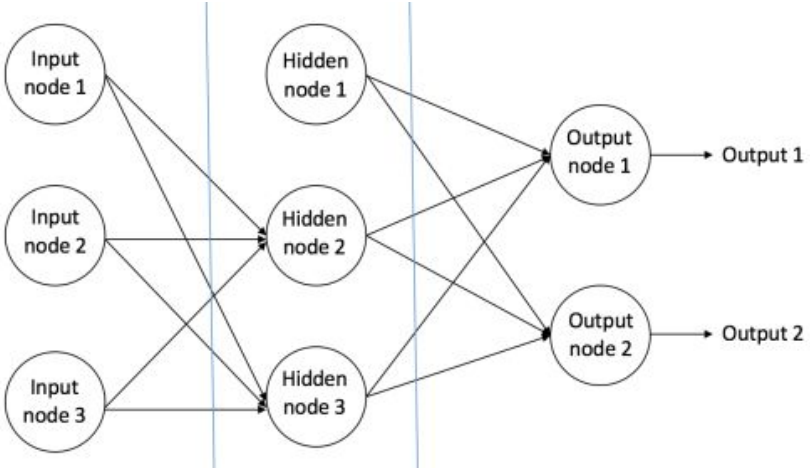
Session 9-10: Introduction to neural networks + Types of neural network + Word embeddings

Sessions 11-13 (Dr. Yukun Lai): Neural network concepts + CNNs, Autoencoders

Session 14 (this): Recap

Neural Networks

Neural networks are based on the interaction of **neurons** (nodes) and **weights** through mathematical functions (known as **activation functions**).



Neural networks are often referred to **deep learning**, especially for large (deep) networks.

Introduction to neural networks:
https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/

Feedforward neural network

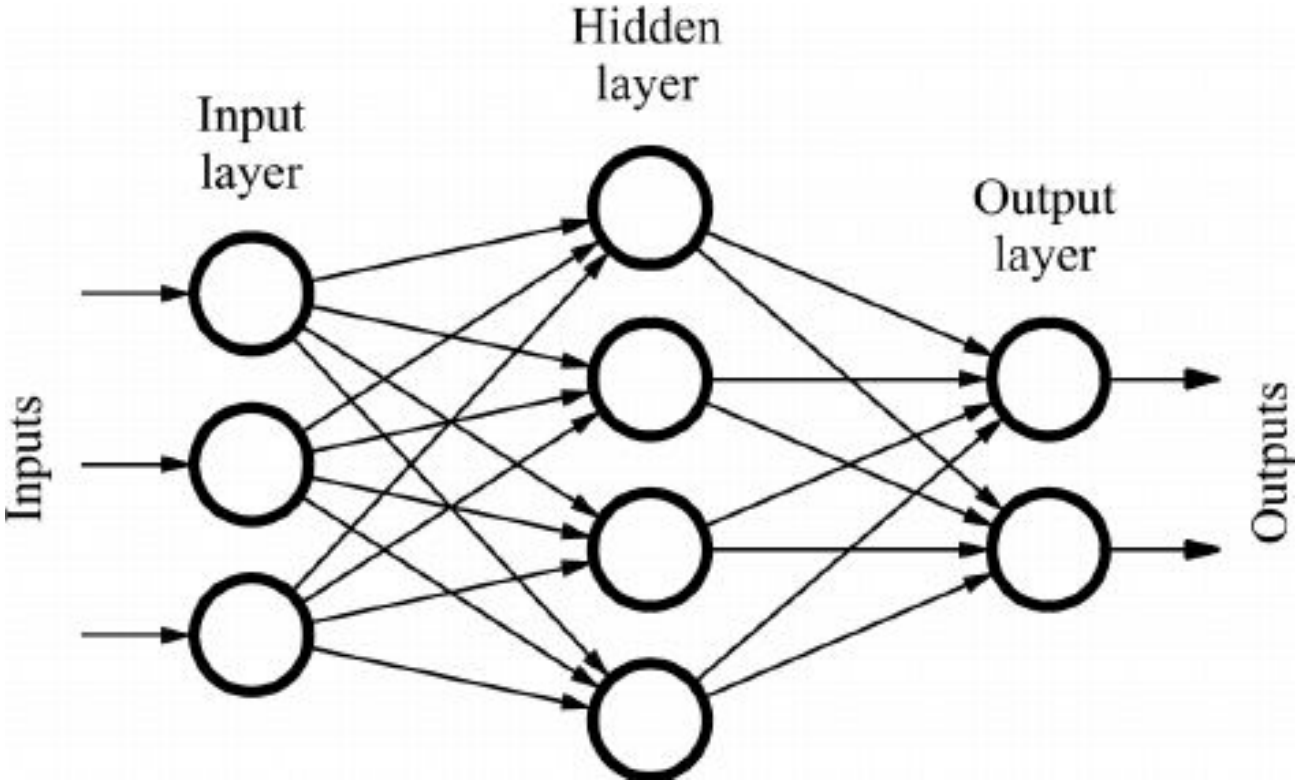


Image credit: *Random TechPark*

Recurrent neural network (LSTM)

→ Highly successful in NLP!

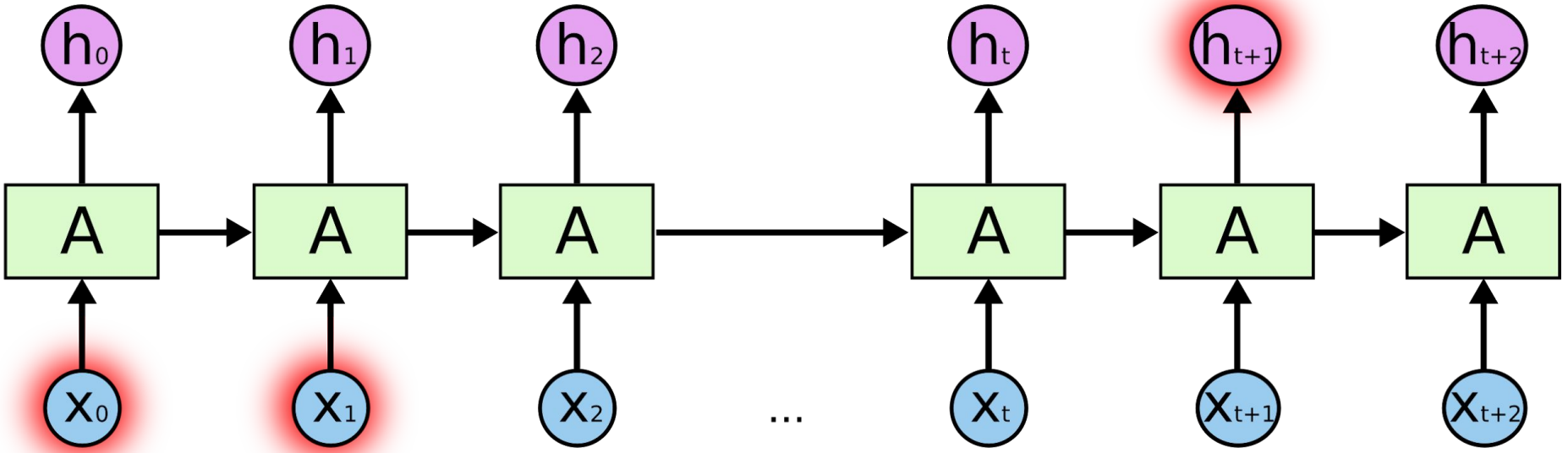


Image credit: *colah's blog*

Convolutional neural network

→ Highly successful in Computer Vision!

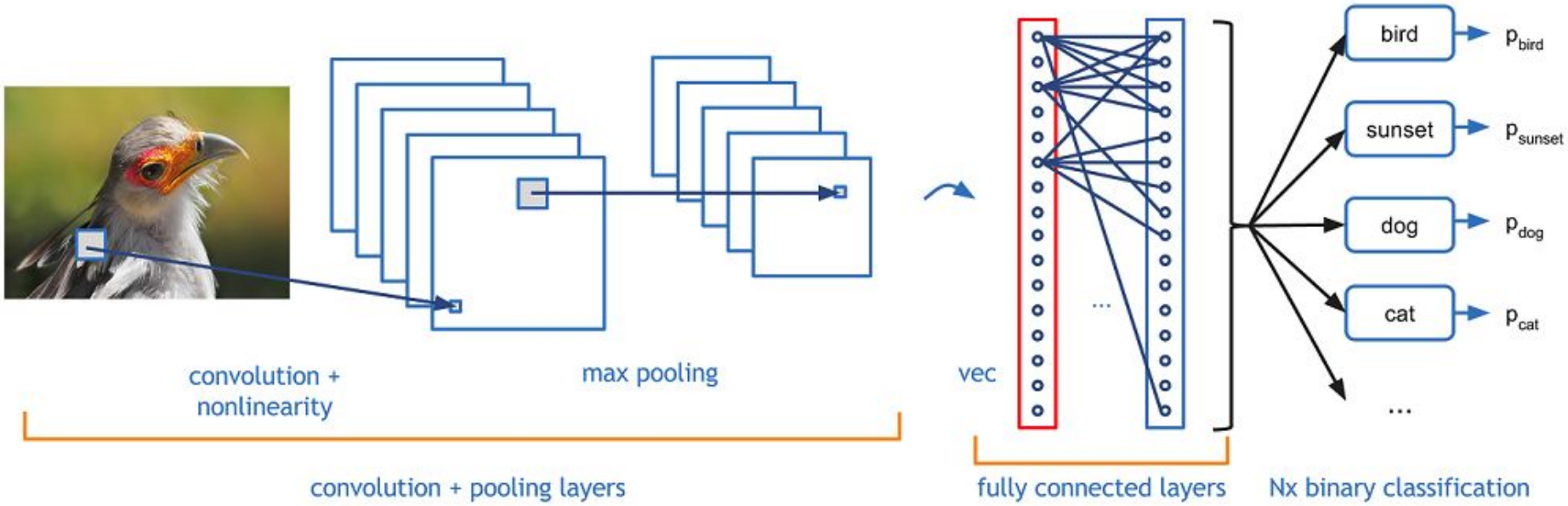


Image credit: *Adit Deshpande*

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

School's private Stack Overflow

https://stackoverflow.com/c/comsc

If technical questions related to the code/implementation, you can post your questions in our private Stack overflow.

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

Guest talks:
Career opportunities - Companies / research

Companies / research

- **Amplyfi:** Unlock the power of the web with AI - business intelligence
  - Online talk in Zoom next week - please contact me by email if interested
- **Hypercascade:** Machine Learning to solve business problems
  - Online talk in Teams next Thursday at 2pm.
- **School of Medicine:** Machine Learning for cancer
  - Video to be shared in Learning Central

AMPLYFI

AMPLYFI is transforming **business intelligence (BI) and research**.

Using its expertise in **deep-web sourcing, artificial intelligence (AI) and information aggregation**, AMPLYFI is providing innovative technologies to businesses and enabling them to make **smarter decisions faster**.

Incorporated in 2015, AMPLYFI is an award-winning company headquartered in Cardiff, UK, that has worked with clients **across all sectors**.

AIRBUS

Bayer

Deloitte.

LLOYDS BANKING GROUP

RBS
The Royal Bank of Scotland

UNIVERSITY OF CAMBRIDGE

BNP PARIBAS FORTIS

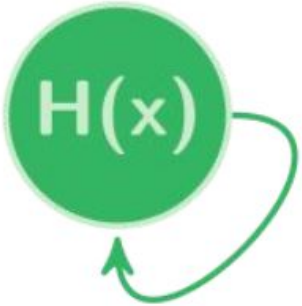
Sumitomo

HARVARD UNIVERSITY

Intro

AMPLYFI Machine Learning R&D is focused on **Information Extraction** from unstructured data. The aim is to gather knowledge latent in text and use it, by itself or in combination with sources of structured data, to ultimately generate insight.

Natural Language Processing and Understanding are the tools that allow automatic Information Extraction, therefore are central to our development.



https://hypercascade.com

Talk: *When Machine Learning meets real-world data*

When: Thursday, 26 March, 2pm (Teams)

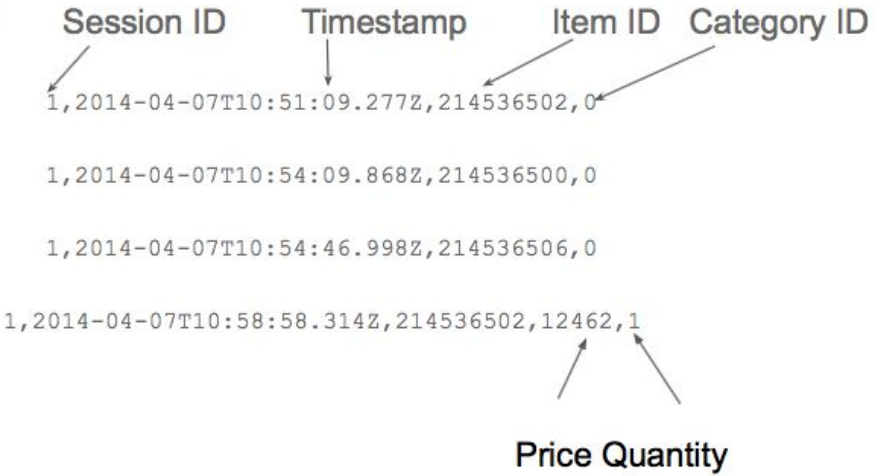
Who: Humphrey Sheil, Founder + PhD Cardiff Uni

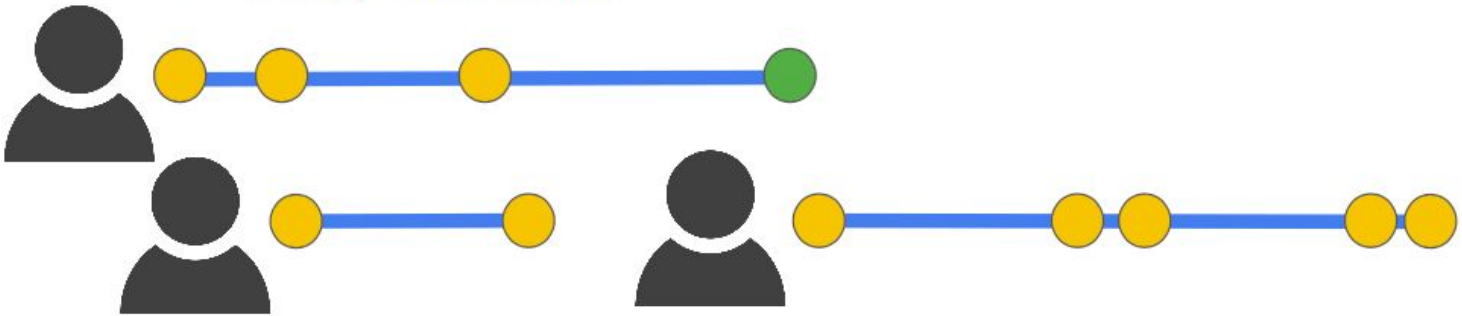
Positions: Summer+year interns, full-time

Cardiff Uni Careers ID: D3Y6E

Areas of research

- Ecommerce event stream prediction
  - Recurrent neural networks
- Semi-structured text
  - Topic modelling (LDA)
  - Word + sentence language models
    - Antonyms, synonyms
  - Content freshness detection (time)
- Tabular data
  - Gradient boosted decision trees
    - Strongly imbalanced data





2

Thank you