CMT307: Applied ML Session 10

Neural network architectures
The case of word embeddings

Jose Camacho Collados

Outline

- Coursework 1
- Neural network architectures
- The case of word embeddings
- Mid-module feedback
- Coursework 2: Group projects

Class: time distribution from **today**

14:10-15:00 -> Lecture

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

15:40-15:55 -> Break (time to move to the group tables)

15:55-17:00 -> Group projects

This time distribution may vary from class to class.

Lectures could include practical activities as well.

Coursework 1

Coursework 1

Marks and feedback will be available in Learning Central **tomorrow**.

Coursework 1: Revision dates

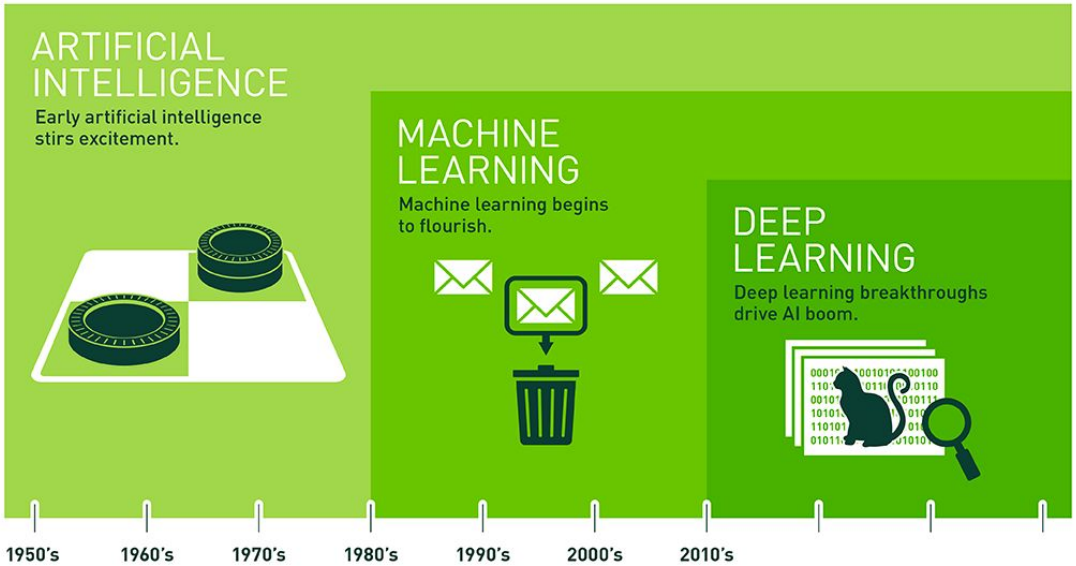
Two options:

- During office hours (**Wednesday, February 12th from 2:30pm to 5pm**).

- Tuesday afternoon, **February 11th**, also possible **if pre-arranged via email**.

Neural network architectures

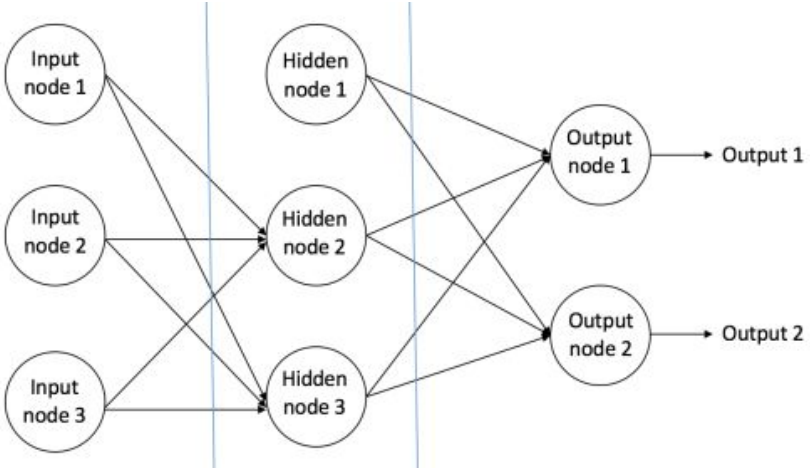
Deep Learning, Machine Learning and AI



Slide credit: *NVidia*

Neural Networks: reminder

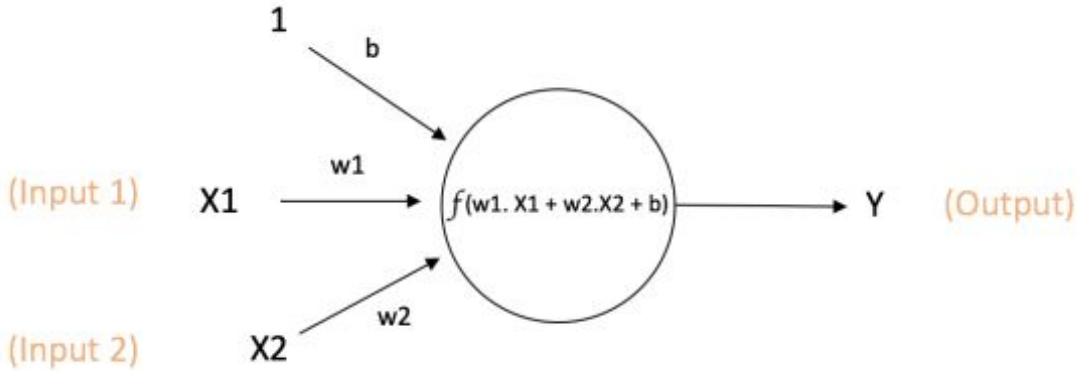
Neural networks are based on the interaction of **neurons** (nodes) and **weights** through mathematical functions (known as **activation functions**).



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/

Perceptron: the smallest neural network



Output: *f(w1.X1+w2.X2+b)*

b is the bias, **X1** and **X2** are the input nodes, **w1** and **w2** are their associated weights and **f** is the activation function. Each iteration, weights are updated through **backpropagation**.

Image credit: *the data science blog*

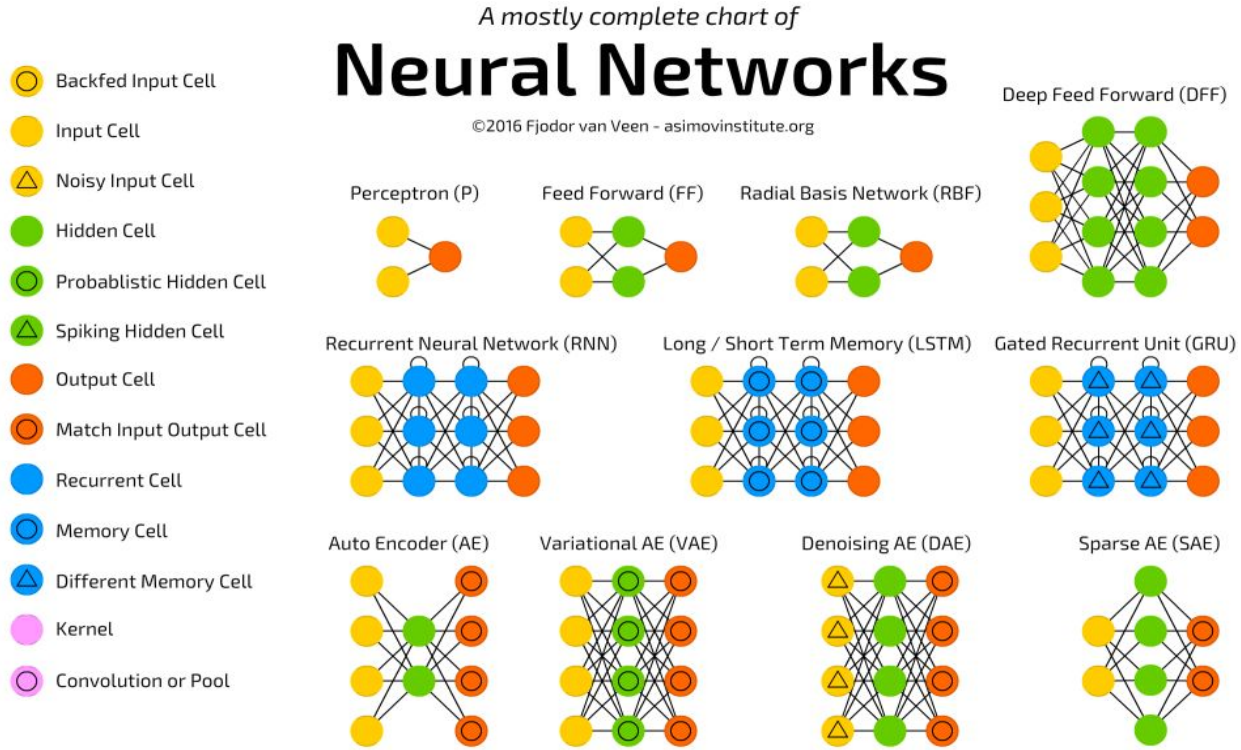
Types of neural network

There are **many types of neural network**, which may be useful for different applications.

Moreover, each type of network has thousands of ramifications and combinations depending on the number of layers, layer size, etc.

More on types of neural networks:
https://www.digitalvidya.com/blog/types-of-neural-networks/

Types of neural network



Feedforward neural network

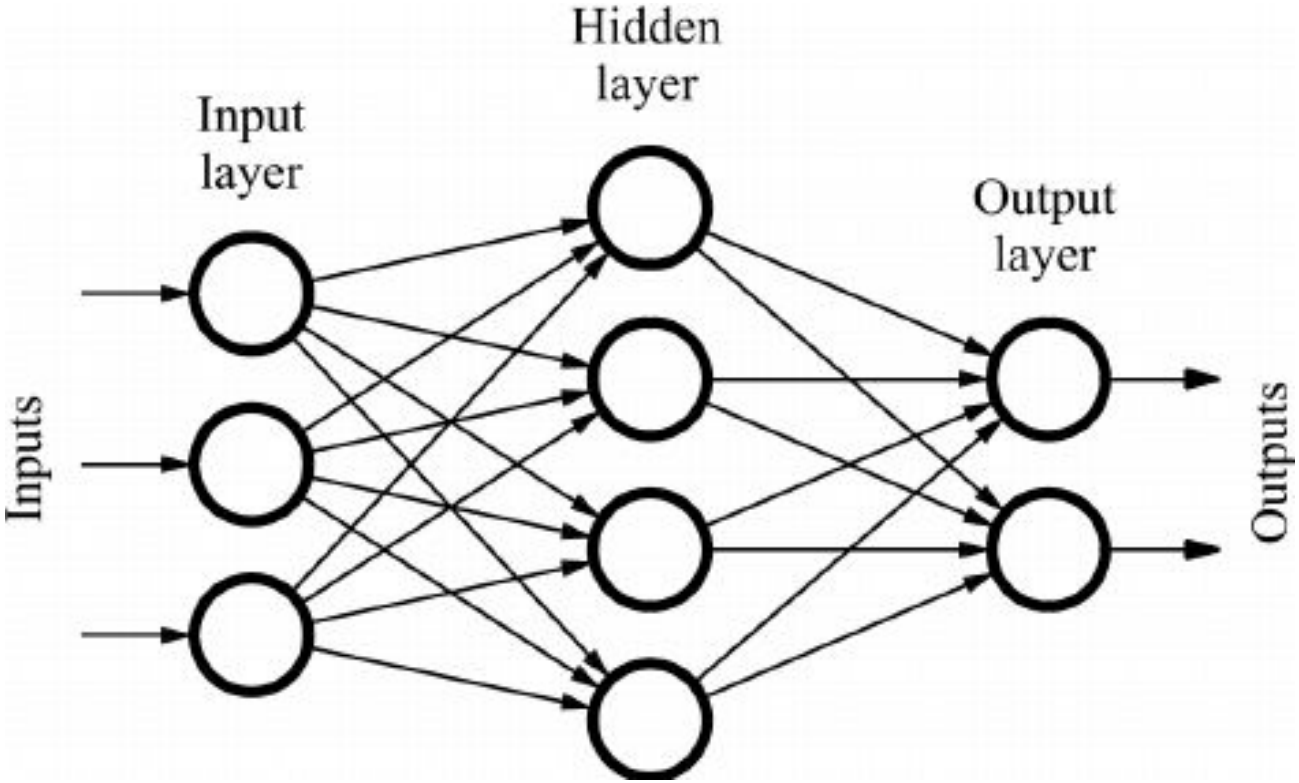


Image credit: *Random TechPark*

Convolutional neural network

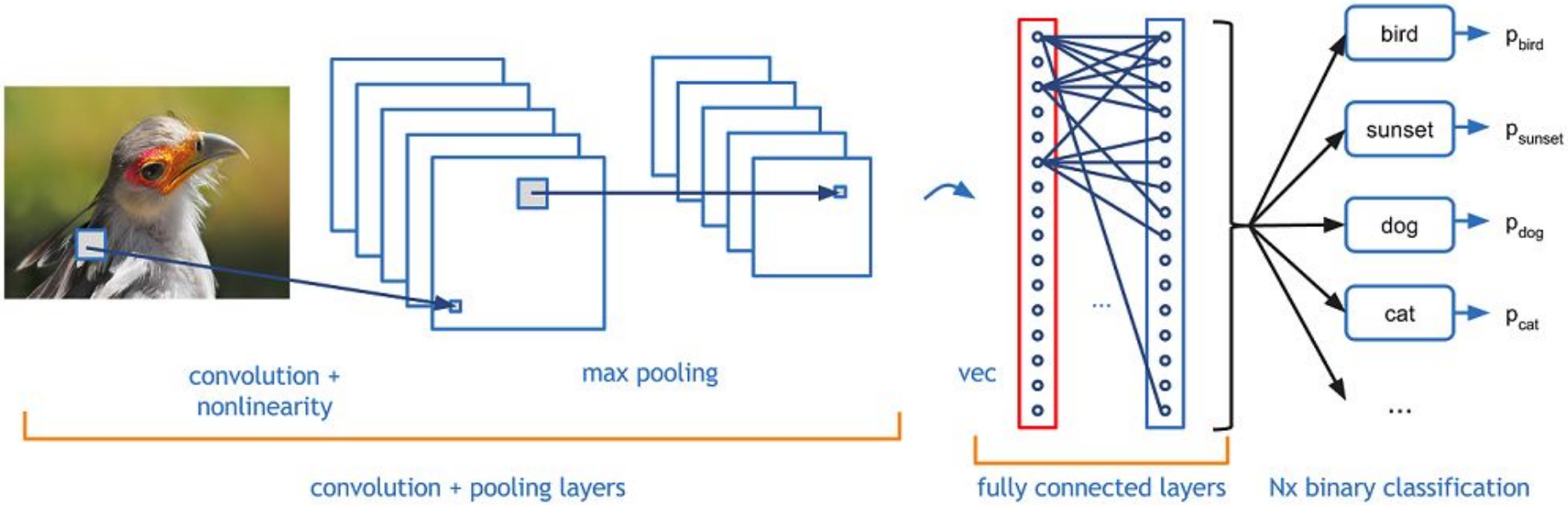


Image credit: *Adit Deshpande*

Recurrent neural network (LSTM)

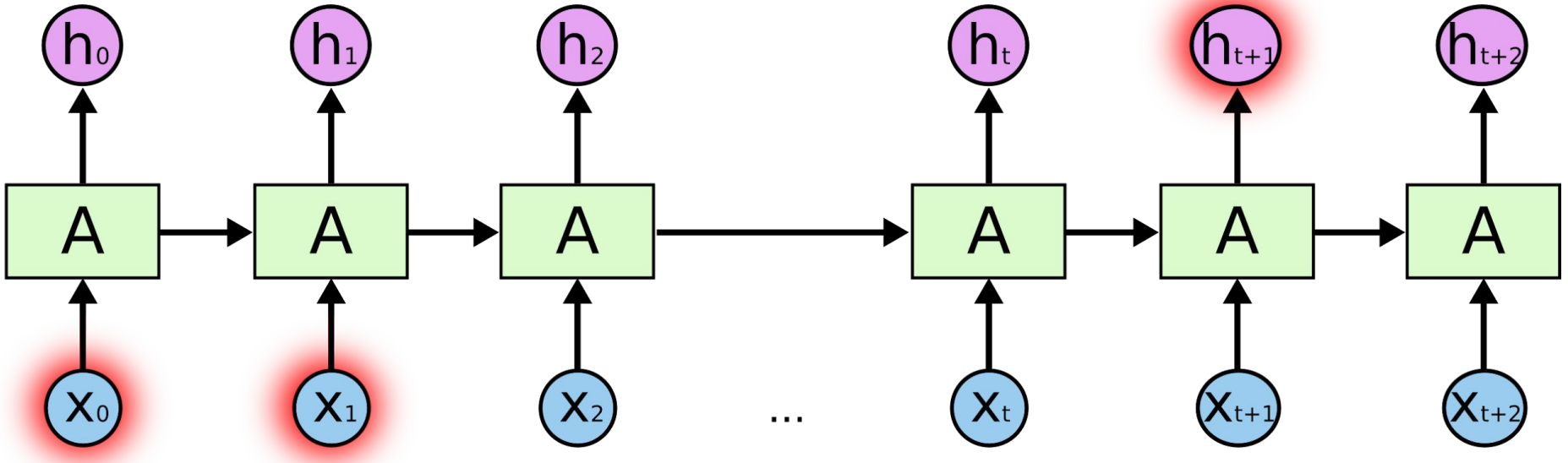


Image credit: *colah's blog*

Word embeddings

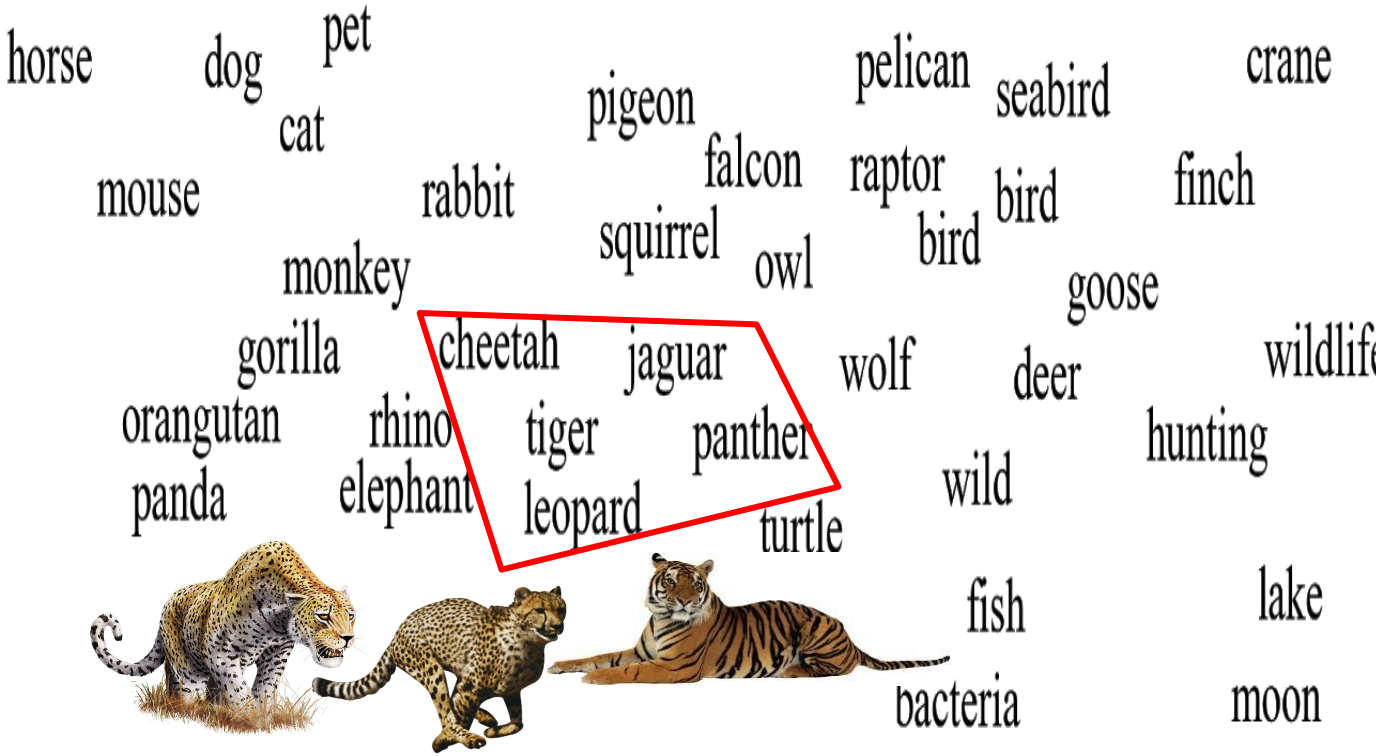
Word embeddings

Word embeddings are **vector spaces** where words are represented as points in the space.

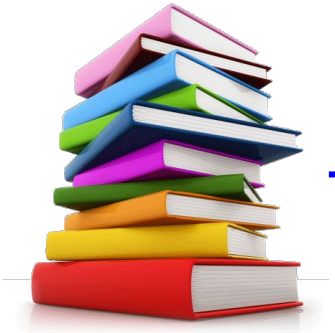
Similar words are represented close in the vector space.

Useful for many Natural Language Processing (**NLP**) applications.

Word embeddings



Word embeddings: How to learn them



... **London** is the **capital** of **UK** …

... Last night I **travelled** from **Cardiff** to **London**.

.

.

.

London

[0.25, 0.32, -0.1 …. 0.1]

Word2Vec (Mikolov et al. 2013)

Word2Vec is one of the pioneers works to learn word embeddings from text corpora.

The architecture is quite simple, a shallow neural network with a single hidden layer.

More about word2vec:
https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa
https://arxiv.org/pdf/1301.3781.pdf (original paper)

Word2Vec: CBOW and Skip-gram

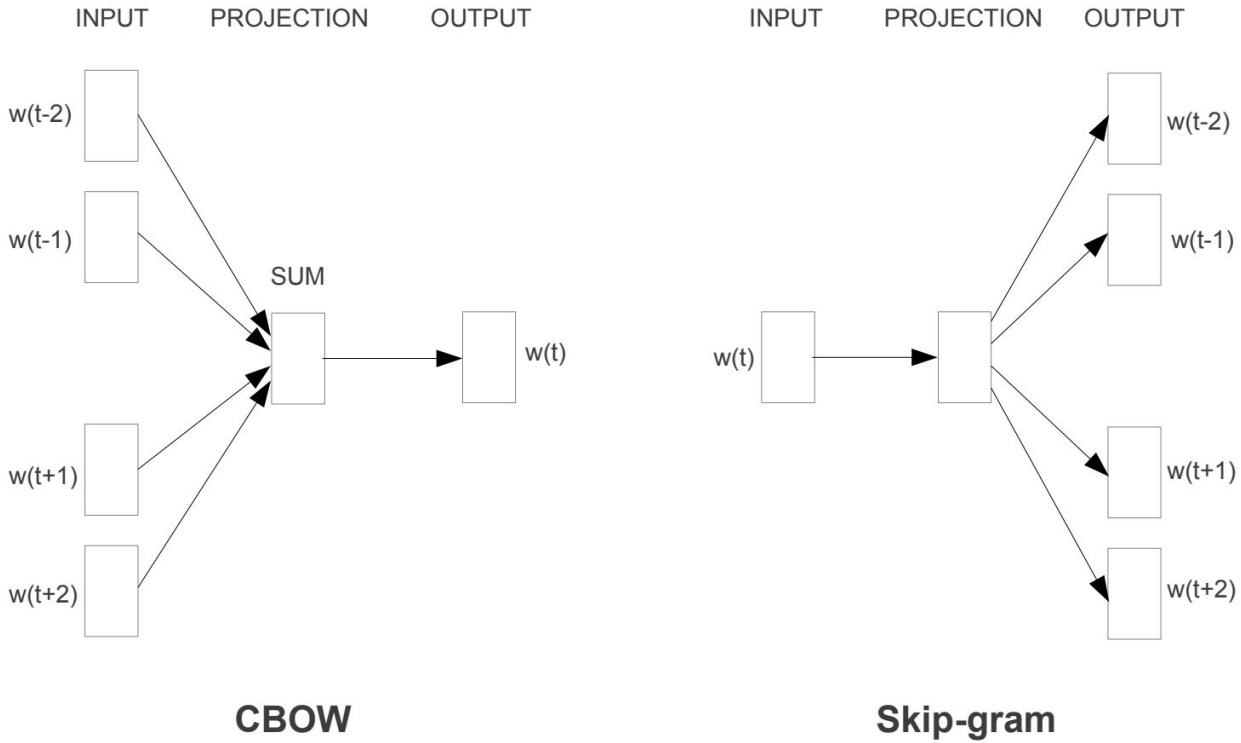
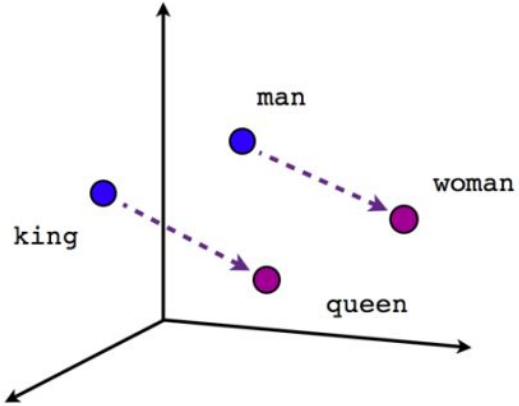
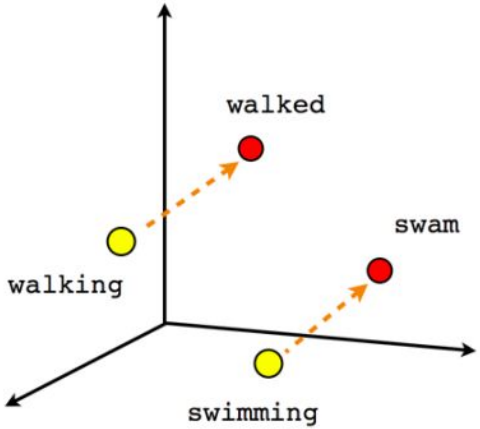


Image credit: *Mikolov et al.*

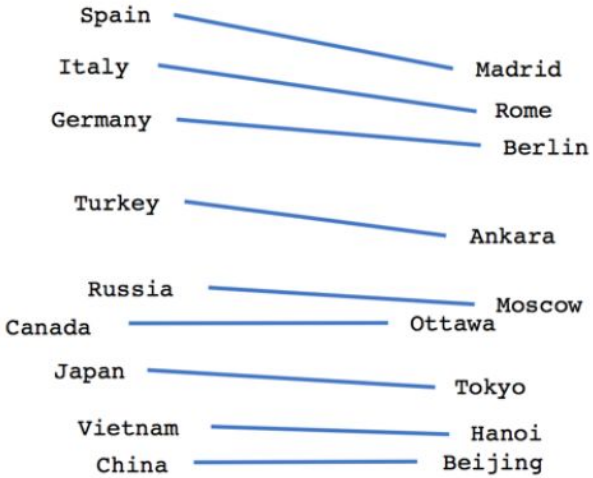
Word2Vec: linguistic regularities



Male-Female



Verb tense



Country-Capital

Image credit: *Mikolov et al.*

Word embeddings as input to neural networks

Word embeddings are often **used as input** in neural network architectures. In many frameworks (e.g. Keras) this initial layer is referred to as **embedding layer**.

With embeddings neural nets gain in **generalization and background knowledge**.

More about word embeddings and neural networks:
www.kdnuggets.com/2018/05/contribution-neural-networks-word-embeddings-natural-language-processing.html

Word embeddings as input to neural networks

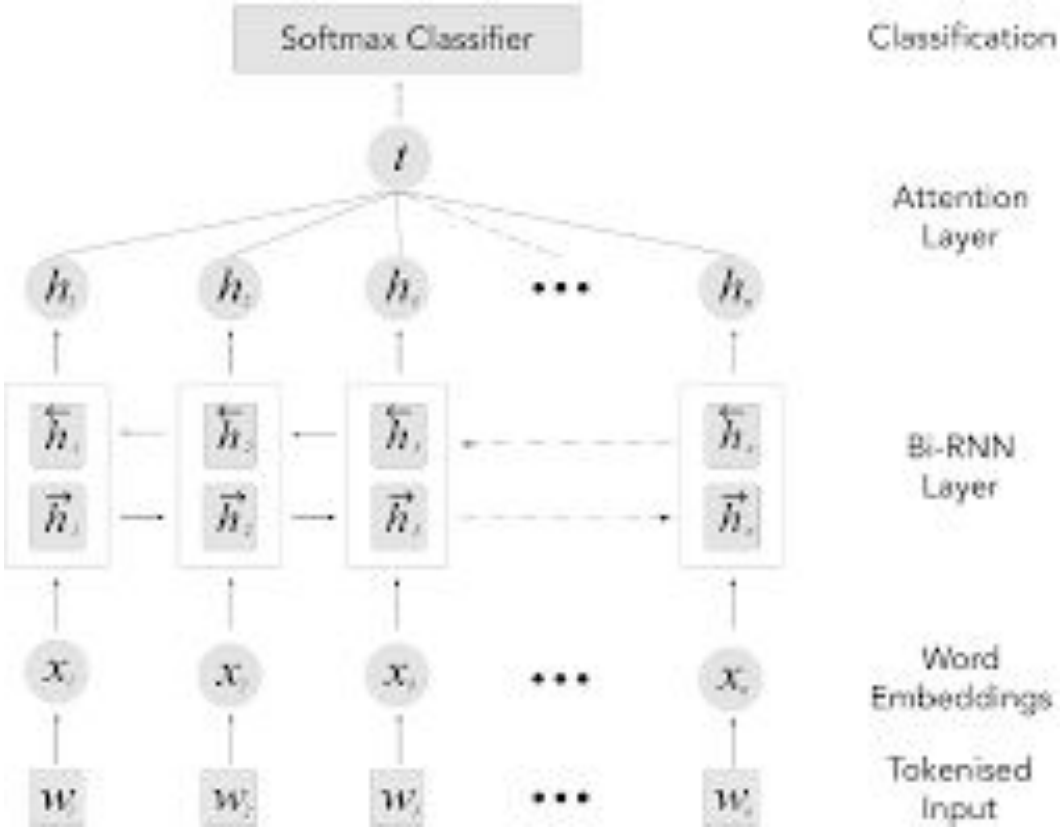


Image credit: *Edo-Osagie et al.*

Word embeddings as input to neural networks

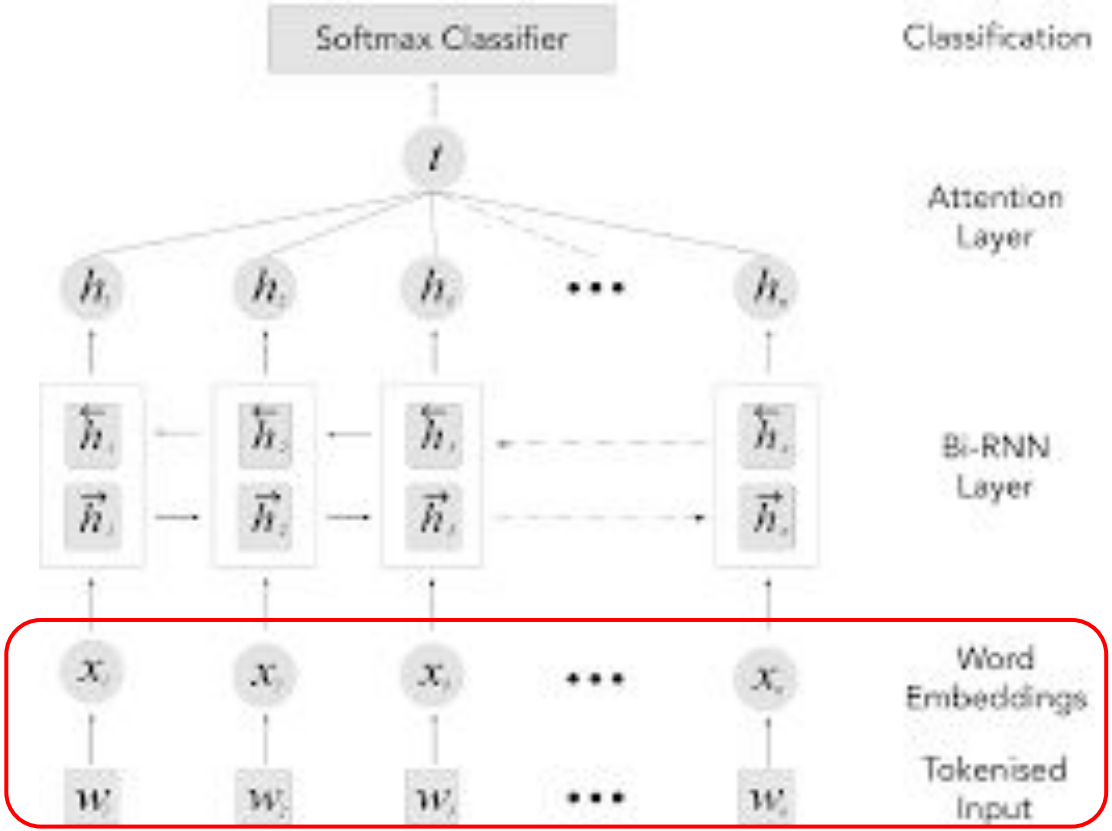


Image credit: *Edo-Osagie et al.*

Contextualized word embeddings

ELMo

Peters et al.
(NAACL 2018)

Based on LSTMs

BERT

Devlin et al.
(NAACL 2019)

Based on Transformers

Contextualized word embeddings

ELMo

Peters et al.
(NAACL 2018)

Based on LSTMs

BERT

Devlin et al.
(NAACL 2019)

More successful nowadays

Based on Transformers

Contextualized word embeddings

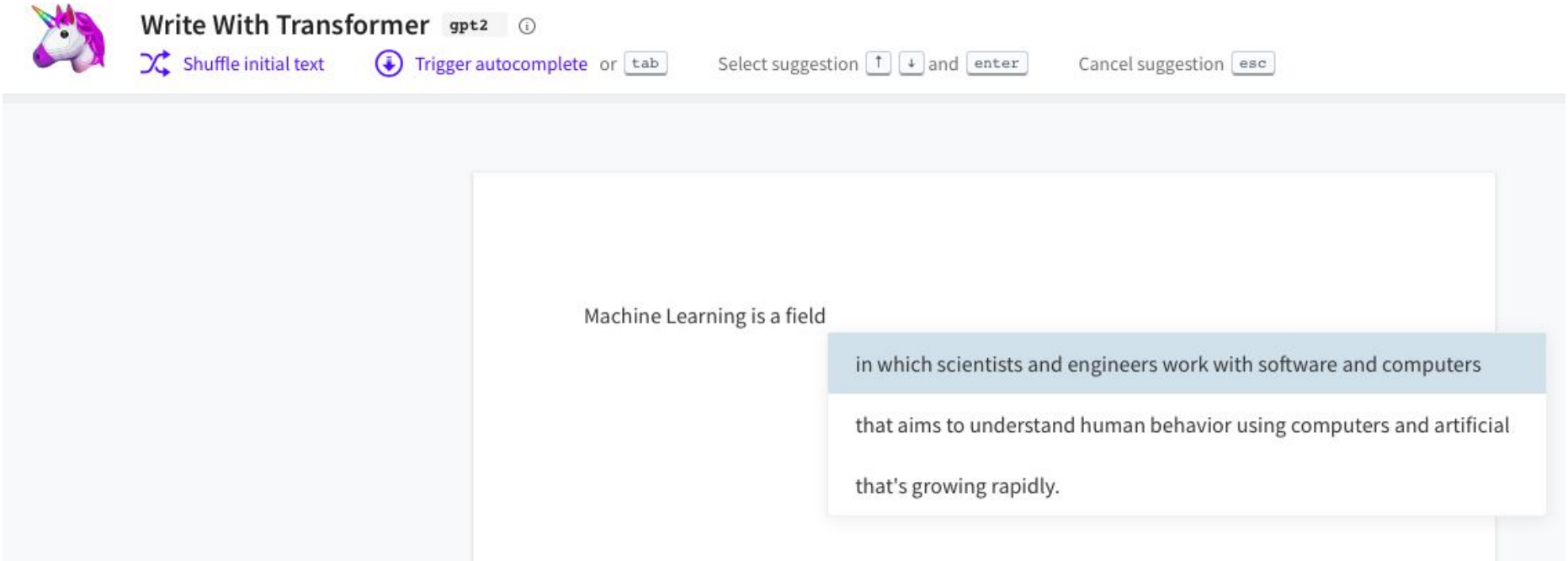
ELMo

BERT

New AI fake text generator may be too dangerous to release, say creators

The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse

Play with transformers / generating text



https://transformer.huggingface.co

Contextualized word embeddings (ELMo/BERT)

As word embeddings, learned by leveraging language models on **massive amounts of text corpora**.

New: each word vector depends on the context. It is dynamic.

Important improvements in many NLP tasks.

Contextualized word embeddings (ELMo/BERT)

0.25, 0.32, -0.1 ….

*He withdrew money from the **bank**.*

0.22, 0.30, -0.08 ….

*The **bank** remained closed yesterday.*

-0.8, 0.01, 0.3 ….

*We found a nice spot by the **bank** of the river.*

Contextualized word embeddings (ELMo/BERT)

0.25, 0.32, -0.1 ….

He withdrew money from the ***bank****.*

0.22, 0.30, -0.08 ….

Similar vectors

The ***bank*** *remained closed yesterday.*

-0.8, 0.01, 0.3 ….

We found a nice spot by the ***bank*** *of the river.*

How to use these embeddings in Python?

Word embeddings:

Easy to use → Gensim

Today's Python Notebook

Pre-trained language models (or contextualized embeddings):

Huggingface transformers

Advanced ↗

Mid-module feedback

Mid-module feedback

Thanks to all who filled the form! 31 answers so far.

Mid-module feedback form: https://bit.ly/2E7o1j0

In the following some more comments are discussed.

Mid-module feedback so far: some comments

*I disliked the lectures in the middle, I mean **all the mathematical and theoretical explanation of ML techniques**.*

*Classes from regression they **were not understandable***

*More focus should be put on **explaining why a particular machine learning method should be chosen** rather than explaining the mathematical theory*

Mid-module feedback so far: some comments

The Stackoverflow page for the coursework discussions did not work for me, I asked around other people on the ***Physics Msc and they also said that they could not access it****.*

Add those from the Physics Msc to the Stackoverflow page*.*

New: *it should work now!*

Mid-module feedback so far: some comments

Using pre-built functions and connecting them together in simple examples feels like cheating

[...]

I think there should definitely be a part of the module dedicated to ***creating as many of our own fundamental functions*** *as possible. If this is beyond the scope of the course, then there should be* ***resources to help us do this in our spare time****.*

Coursework 2: Group projects

Group projects

In this semester the coursework will consist of **group projects**:

- **Goal:** Design and implement machine learning models for solving a problem.
- **Tasks:** Write a group report, give a presentation. Write an individual report.

All projects and corresponding datasets: http://tiny.cc/l3ccjz

Coursework 2: Instructions

Draft of the coursework assignment is now available here:
http://tiny.cc/1anmjz

The **final version** of the assignment will be available **in Learning Central** next week.

Group project allocation

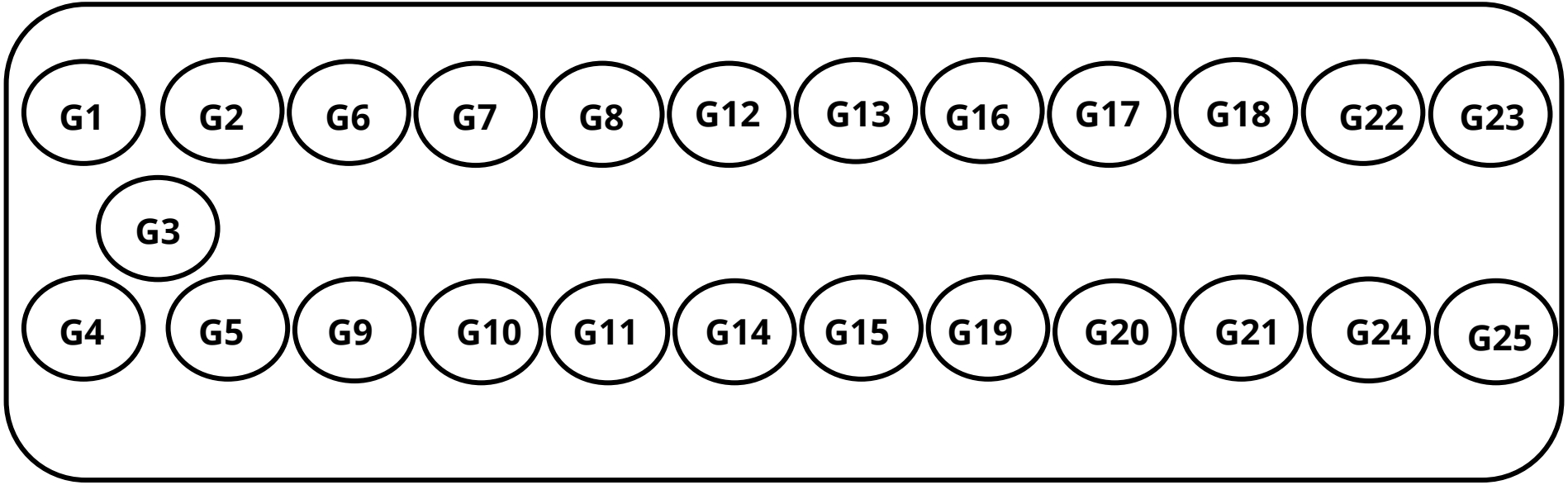
Available here:

http://tiny.cc/84wmjz

All **preferences have been respected**.

After the break, each **group will be allocated a table**.

Table allocation



Group project: What to do now?

Download the datasets. Get to know each other, exchange emails/contact details. Then, decide the main task for each member of the group:

- **Descriptive analysis of the dataset + Error analysis**
- **Preprocessing + Literature review**
- **Implementation + Results**

(At least two students per task)

It could be also useful to decide the **main contact person** of the group for communication with supervisors/instructors.

Group projects: Milestones

- **February 20:** Initial form with allocation of tasks within group members.
- **March 4:** Mid-report (1 page maximum) with progress and first implementation of a machine learning model.
- **April 20:** End of experiments (ideally earlier) - **no more models to run from this date** (time to write the report and finish the analysis)

Group projects: Communication with the supervisor

In most lectures on Thursdays **your supervisor will be present during the last hour (from 4pm to 5pm)**. This will be a moment to **show your progress** to your supervisor and also ask for advice/doubts.

The time will be limited, so **groups should prepare in advance** for this meeting.

Supervisors are also **available via email** for group communication (contact details in Learning Central).

Questions?

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.

Hands on!

Python notebook with exercises
about **word embeddings in gensim** available at Learning Central.

We are available for questions during this time