

NLP Future Me Lab Sheet

Exercise 1 – Word Embeddings - Slides 1-18

Part A – Word2Vec Demo

Navigate to: <https://remykarem.github.io/word2vec-demo/>

This site allows you to visualise the relationship between words in a source text. First try it with the given example (the lyrics to hey there Delilah) and see if you can spot any interesting factors in the results.

Next, try taking one of the following texts (N.b., just copy a few paragraphs, otherwise we will kill the service) and copy-pasting it into the first box on the website. Follow through the prompts on the website to see the results.

- 1) Particle physics Wikipedia: https://en.wikipedia.org/wiki/Particle_physics
- 2) Moby Dick (search for chapter 1. Loomings. For the start):
<https://www.gutenberg.org/cache/epub/2701/pg2701.txt>
- 3) News Text: <https://www.bbc.co.uk/news/articles/cvgp6y42rl3o>
- 4) Find your own text from the web and try it out!

Part B – Word Embeddings Playground

Navigate to: <https://wordembeddings.xyz/analogy>

Try out some analogies on the ‘Analogy’ tab. Some easy ones to get you started:

Man is to **King** as **Woman** is to ?

London is to **England** as **Paris** is to ?

Will is to **Would** as **Can** is to ?

Rain is to **Manchester** as **Sunshine** is to ?

Try out some of your own too.

You may also wish to try the ‘Groups’ tab. A good example group to start with would be: man, woman, boy, girl, king, queen, prince, princess

Exercise 2 – Text Classification – Slides 19-26

Download the jupyter notebook here:

https://github.com/mattshardlow/NLP_FutureMe/blob/main/sequence_classification.ipynb

and upload it to Google Colab:

<https://colab.research.google.com/>

Follow the instructions in each cell and experiment with the results. By the end of the notebook you will have trained a version of Bert for text classification (in this case to classify reviews of films from IMDB).

Exercise 3 – Tokenisation – Slides 27-41

Try the following examples in the tokeniser at:

<https://platform.openai.com/tokenizer>

What do you notice in each case? What does this tell you about the sequence that an LLM is able to interpret?

Example 1:

“Begone, foul dwimmerlaik, lord of carrion! Leave the dead in peace!”

A cold voice answered: 'Come not between the Nazgûl and his prey! Or he will not slay thee in thy turn. He will bear thee away to the houses of lamentation, beyond all darkness, where thy flesh shall be devoured, and thy shrivelled mind be left naked to the Lidless Eye.'

Example 2:

Castell Coch is a 19th-century Gothic Revival castle built above the village of Tongwynlais in South Wales. The first castle on the site was built by the Normans after 1081 to protect the newly conquered town of Cardiff. The castle's earth motte was reused by Gilbert de Clare as the basis for a new stone fortification, built between 1267 and 1277. John Crichton-Stuart, 3rd Marquess of Bute, inherited the castle ruins in 1848. One of Britain's wealthiest men, he employed the architect William Burges to reconstruct the castle as a summer residence.

Example 3:

Aaren,Aarika,Abagael,Abagail,Abbe,Abbey,Abbi,Abbie,Abby,Abbye,Abigael,Abigail,Abigale,Abra,Ada,Adah,Adaline,Adan,Adara,Adda,Addi,Addia,Addie,Addy,Adel,Adela,Adelaida,Adelaide,Adele,Adelheid,Adelice,Adelina,Adelind,Adeline,Adella,Adelle,Adena,Adey,Adi,Adiana,Adina,Adora,Adore,Adoree,Adorne,Adrea,Adria,Adriaens,Adrian,Adriana,Adriane,Adrianna,Adrienne,Adriana,Adrienne,Aeriel,Aeriela,Aeriell,Afton,Ag,Agace,Agata,Agatha,Agathe,Aggi,Aggie,Aggy,Agna,Agnella,Agnes,Agnese,Agnesse,Agneta,Agnola,Agretha,Aida,Aidan,Aigneis,Aila,Aile,Ailee,Aileen,Ailene,Ailey,Aili,Ailina,Ailis,Ailsun,Ailyn,Aime,Aimee,Aimil,Aindrea,Ainslee,Ainsley,Ainslie,Ajay,Alaine,Alameda,Alana,Alanah,Alane,Alanna,Alayne,Alberta,Albertina,Albertine,Albina,Alecia,Aleda,Aleece,Aleen,Alejandra,Alejandrina,Alena,Alene,Alessandra,Aleta,Alethea,Alex,Alexa,Alexandra,Alexandrina,Alexi,Alexia,Alexina,Alexine,Alexis,Alfi,Alfie,Alfreda,Alfy,Alia,Alicia,Alice,Alicea,Alicia,Alida,Alidia,Alie,Alika,Alikee,Alina,Aline,Alis,Alisa,Alisha,Alison,Alissa,Alisun,Alix,Aliza,

Alla,Alleen,Allegra,Allene,Alli,Allianora,Allie,Allina,Allis,Allison,Allissa,Allix,Allsun,Allx,Ally,Allyce,Allyn,Allys,Allyson,Alma,Almeda,Almeria,Almeta,Almira,Almire,Aloise,Aloisia,Aloysia,Alta,Althea,Alvera,Alverta,Alvina,Alvinia,Alvira,Alyce,Alyda,Alys,Alysa,Alyse,Alysia,Alyson,Alyss,Alyssa,Amabel,Amabelle,Amalea

Exercise 4 – Text Generation – Slides 42-96

Download the jupyter notebook here:

https://github.com/mattshardlow/NLP_FutureMe/blob/main/Text_Generation_with_GPT.ipynb

and upload it to Google Colab:

<https://colab.research.google.com/>

Follow the instructions in each cell and experiment with the results. By the end of the notebook you will have a version of several LLMs including GPT-2 and an instruction fine-tuned version of Bloom that you can experiment with. Try adjusting parameters such as the temperature and the top_p value.