UNIVERSIDAD DE ANTIOQUIA
1803

Lab III - Advanced Topics

Machine Learning II

Prerequisites

Python

Python

- Data structures (properties of lists, tuples, dicts, built-in modules…)
- Classes
- Packages and modules

NumPy

- Arrays
- Inner product
- Vector - Matrix product
- Distances

Workshop III

1. In your own words, describe what vector embeddings are and what they are useful for.

2. What do you think is the best distance criterion to estimate how far two embeddings (vectors) are from each other? Why?

3. Let us build a Q&A (question answering) system! 😀For this, consider the following steps:
    a. Pick whatever text you like, in the order of 20+ paragraphs.
    b. Split that text into meaningful chunks/pieces.
    c. Implement the embedding generation logic. Which tools and approaches would help you generate them easily and high-level?
    d. For every question asked by the user, return a sorted list of the N chunks/pieces in your text that relate the most to the question. Do results make sense?

4. What do you think that could make these types of systems more robust in terms of semantics and functionality?
5. Bonus points if deployed on a local or cloud server.

Useful Resources

- https://huggingface.co/
- https://www.slideshare.net/perone/word-embeddings-introduction
-