

DATA.ML.100 Introduction to Pattern Recognition and Machine Learning  
TAU Computing Sciences  
Exercise - Week 3 *Nearest neighbor search*

Be prepared for the exercise sessions (watch the demo lecture). You may ask TAs to help if you cannot make your program to work, but don't expect them to show you how to start from the scratch.

1. **Introduction**

Using machine learning techniques, words of any language can be embedded into a high dimensional Euclidean space where semantically similar words are close to each other.

You are provided a file that contains 400,000 English terms and their 50-dimensional embedding vectors.

[https://tuni-my.sharepoint.com/:t:/g/personal/joni\\_kamarainen\\_tuni\\_fi/ER02557-gp5JrkElvQITa0YB-TjrA?e=bhio7j](https://tuni-my.sharepoint.com/:t:/g/personal/joni_kamarainen_tuni_fi/ER02557-gp5JrkElvQITa0YB-TjrA?e=bhio7j)

You are also provided a code that loads the words and their vectors.

2. **Search similar words** (20 points)

For any input word, return the three (3) most similar words (the most similar should be the input word itself). Give results at least to:

- king
- europe
- frog

3. **Search analogy** (30 points)

Another interesting task is the search of analogy, for example, “king is to queen as prince is to X” - what would be X? In the word embedding space this can be done using the difference vectors. If  $x$  is the king word vector,  $y$  is the queen word vector and  $z$  is the prince word vector, then the Euclidean version of analogy is  $z = z + (y - x)$ , that is, the vector from king to queen is added to prince to obtain vector  $z$  that is relatively in the same location as queen is from king. The corresponding word must then be sought using the nearest neighbor search.

For each analogy, return two (2) best matches, for the following combinations

- king-queen-prince
- finland-helsinki-china
- love-kiss-hate

Return the following items:

- Python code: <surname>\_word\_search.py
- A full desktop screenshot that includes the terminal window where code is run:  
<surname>\_word\_search\_screenshot.png
- Python code: <surname>\_word\_analogy.py
- A full desktop screenshot that includes the terminal window where code is run:  
<surname>\_word\_analogy\_screenshot.png