

Introduction to LLM

Practice Session 2

CL/NLP Basics

The Focus of Practice Session

- We revisit the concepts from the lecture, but in a practical / hands-on way.
- For most concepts, we start with a tiny, simple implementation so you see how it works inside.
- Then we look at how to use the same idea with reliable library functions.
 - This is how you would actually do it in real code projects.
- This means you don't only learn how to call it, but you also understand what it's doing internally.
- In the practice session, the goal is:
 - to follow the logic and flow of the code, and
 - to be able to complete missing parts of the tasks in a reasonable, logical way.

Coding in Exam

- We focus on understanding, not memorization.
- You will not lose points for forgetting exact library function names.
- You will get points for writing the correct idea / logic.
- We care that you can show what the algorithm does, not just code syntax.
- Pseudo-code is OK if it shows the correct steps clearly.
- Syntax mistakes are not critical.
- Big logic mistakes are.

Correct implementation (valid code, logically correct)

```
cosine_similarity(a, b):
```

```
    dot = 0.0
```

```
    for i from 0 to length(a)-1:
```

```
        dot = dot + a[i] * b[i]
```

```
    mag_a = 0.0
```

```
    mag_b = 0.0
```

```
    for i from 0 to length(a)-1:
```

```
        mag_a = mag_a + a[i] * a[i]
```

```
        mag_b = mag_b + b[i] * b[i]
```

```
    mag_a = sqrt(mag_a)
```

```
    mag_b = sqrt(mag_b)
```

```
    if mag_a == 0 OR mag_b == 0:
```

```
        return 0.0
```

```
    return dot / (mag_a * mag_b)
```



$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Pseudo-code (not real code, but still logically correct)

```
cosine_similarity_pseudo(a, b):
```

```
    dot = Σ over i of (a[i] * b[i])
```

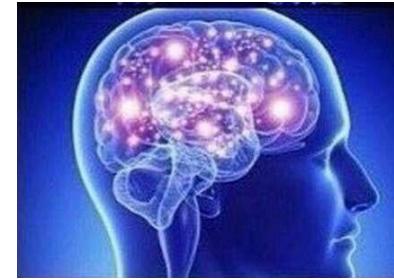
```
    len_a = sqrt( Σ over i of (a[i]^2) )
```

```
    len_b = sqrt( Σ over i of (b[i]^2) )
```

```
    if len_a == 0 OR len_b == 0:
```

```
        return 0
```

```
    return dot / (len_a * len_b)
```

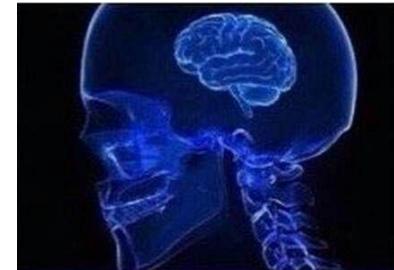


$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Wrong logic (looks like code, but is NOT cosine similarity)

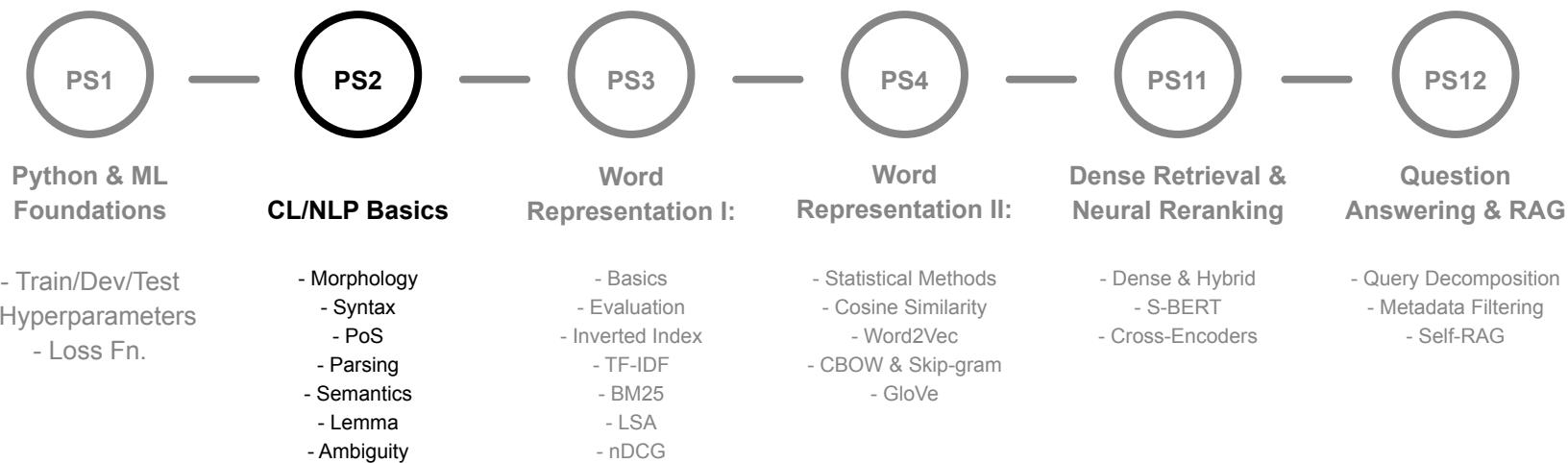
```
cosine_similarity_wrong(a, b):
```

```
dot = 0.0  
  
sum_a = 0.0  
  
sum_b = 0.0  
  
for i from 0 to length(a)-1:  
    dot = dot + a[i] * b[i]  
  
    sum_a = sum_a + a[i]  
  
    sum_b = sum_b + b[i]  
  
return dot / (sum_a * sum_b)
```



$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Timeline



PS2: Colab Notebook (Available on Moodle)



The screenshot shows a Google Colab interface with the following details:

- Title:** Copy of Practice_Session_01_Student_Exercises.ipynb
- Toolbar:** File, Edit, View, Insert, Runtime, Tools, Help
- Search Bar:** Commands
- Code Editor:** A large text area containing the content of the notebook.
- Toolbar Buttons:** + Code, + Text, Run all, Copy to Drive (highlighted with a dashed box and arrow).
- Table of Contents:**
 - PS 01: Introduction (Python and ML Foundations)
 - Learning Objectives
- Description:** By the end of this practice session, you will be able to:
- Objectives List:**
 1. Know Python basics: variables, data types, operators, and control structures
 2. Manipulate strings effectively: indexing, slicing, and built-in string methods
 3. Work with data structures: lists, dictionaries, and their operations
 4. Handle file I/O: reading/writing text files and JSON data
 5. Use essential libraries: NumPy for numerical computing, Pandas for data manipulation
 6. Apply object-oriented programming: classes, methods, and type hints
 7. Implement basic ML workflows: data splitting, model training with PyTorch

- https://colab.research.google.com/drive/1EFXK8CyUVjq2n7Bx1_QOCG_6b3Jbxqib