

Lập trình Python căn bản - Day 8

Ngày 18 tháng 6 năm 2024

Ngày thực hiện:	18/06/2024
Người thực hiện:	Đinh Thị Tâm
Nguồn:	AIO2024 - Day 8
Nguồn dữ liệu (nếu có):	Link of Data Sources Day 8
Từ khóa:	Tuple - Cosine Similarity
Người tóm tắt:	Đinh Thị Tâm

1. Mô tả

Cosine Similarity là một phương pháp đo lường mức độ tương đồng giữa hai vectơ trong không gian đa chiều, đặc biệt hữu ích trong các trường hợp có nhiều biến (high-dimensional spaces). Được tính bằng công thức:

$$\text{CosineSimilarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

Trong đó:

- $\text{CosineSimilarity}(A, B)$: Tên của hàm tính toán Cosine Similarity giữa hai vectơ A và B.
- A và B: Hai vectơ cần so sánh.
- $\|A\|$: Độ lớn (length) của vectơ A.
- $\|B\|$: Độ lớn (length) của vectơ B.

2. Code

```
1 import math
2 # Cosine Similarity
3
4
5 def CosineSimilarity(A, B):
6     # tính tích vô thương
7     length = len(A)
8     tmp = 0
9     length_A = 0
10    length_B = 0
11    for idx in range(length):
12        # tính A.B
13        tmp += A[idx]*B[idx]
14        # tính độ lớn của A, B
15        length_A += A[idx]**2
16        length_B += B[idx]**2
17    length_A = math.sqrt(length_A)
```

```
18     length_B = math.sqrt(length_B)
19     return tmp/(length_A*length_B)
20
21
22 # main
23 vector_a = [1, 2]
24 vector_b = [4, 5]
25 print(f'{CosineSimilarity(vector_a, vector_b)}')
```

3. Kết quả thực thi

```
1 0.9778024140774094
```