

笔记 22-9-29

李肖

2022 年 9 月 29 日

1 度量向量的相似度

问题：已知有 3 个向量，代表了评分系统中 3 个人的评分数据，其中 A[0] 是对第一个电影的评分，A[1] 是对第二个电影的评分……

$$A = [2, 4, 1, 3, 5]$$

$$B = [1, 2, 3, 4, 5]$$

$$C = [3, 4, 1, 5, 2]$$

求这三个向量的相似度

方法 1:

欧几里得距离

$$dis(A, B) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots (a_n - b_n)^2}$$

方法 2:

求逆序对，毕竟这里的评分相当于对这 5 个电影的排名，如果这 5 个电影在两个人的心目中的排名数据接近那么这两个人的喜好就很一样。

1.1 $O(n^2)$ 求逆序对

```
1 arr = [3, 4, 1, 5, 2]
2 l = len(arr)
3 cnt = 0
4 for i in range(l):
```

```
5     for j in range(i + 1, l):
6         if arr[j] < arr[i]:
7             cnt += 1
8     print(cnt)
```

1.2 $O(n * \log(n))$ 求逆序对

TIPS: 使用归并排序