

CU6051NI - Artificial Intelligence

Lab work – 8

In this lab work, you will be working on building a **recommendation system** using **collaborative filtering**. Given below is a table of ratings that users have given to certain books. The ratings are in a range of 1-10.

| | Harry Potter | Start With Why | Tom Sawyer | Superman | Rebecca | Hamlet |
|--------|--------------|----------------|------------|----------|---------|--------|
| John | 5 | 7 | 6 | 7 | 5 | 6 |
| Jeff | 5 | 7 | 3 | 9 | 7 | 6 |
| Andrei | 7 | 7 | 6 | 8 | 5 | 9 |
| Mike | 6 | 8 | 4 | 6 | 4 | 6 |
| Mark | 6 | 8 | 9 | 9 | 7 | 6 |
| Jason | | 9 | | 7 | 5 | |

Using the collaborative filtering algorithm, predict the ratings of Jason for the books that he has not rated.

You can use the example discussed in the lecture as reference.

For finding the similarity between users, use the **Euclidean Distance** and see what kind of results you get.

Then repeat the same process again by using **Pearson Correlation Coefficient** to calculate the similarity between users. The formula for calculating the Pearson correlation coefficient is as follows.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Here you can think of x and y as points in a n -dimensional space. \bar{x} and \bar{y} are the respective mean values.

For example, if 2 people from the above table are to be represented in a 3-dimensional space using the rating they have given,

Jason = [9,7,5]

John = [7,7,5]

The following code can be used to calculate the Pearson correlation coefficient between the two.

```
import numpy as np

def pearson(x,y):

    a = sum((x-np.mean(x))*(y-np.mean(y)))

    b = pow(sum(pow(x-np.mean(x),2)),0.5)*pow(sum(pow(y-np.mean(y),2)),0.5)

    return a/b


jason = np.array([9,7,5])
john = np.array([7,7,5])
print(pearson(jason,john))
```

The range of values for the Pearson correlation coefficient is -1 to 1. The closer the value to 1, the more similar the people are and the closer the value to -1 the more dissimilar the people.

Repeat the collaborative filtering algorithm by using the Pearson correlation coefficient and see what kind of results you get.

You are free to do this assignment using any tool you like. You can use python to code everything up or you can do all calculations by hand too. MS Excel is also a good tool that you can experiment with.