# MDM-2021 Assignment 2

Jeyhun Yagublu (SN:1011331 )

October 24, 2021

## 1 Task 4

In this task, you should study an alternative internal validation index, $\tau$ .
Use only the spiral.txt data from the previous task.

### 1.1 A)

**Task**: Compute $\tau$ for the previously obtained clusterings by K-means and
spectral clustering. Which clustering is now the best one?
**Answer**: I have written my own function for finding the tao index value:

```python
def Gaussian_similarity_f(x_i,x_j, sigma) :
    distance=np.linalg.norm(x_i - x_j)**2
    return math.exp((-distance)/(2*(sigma**2)))

def c_i_j (labels,i,j) :
    if (labels[i]==labels[j]):
        return 1
    else:
        return 0

def tao_index (data , labels,sigma) :
    tao=0;
    for i in range (0,len(labels)):
        nominator=0 ;
        denominator=0
        for j in range (0,len(labels)) :
            if(j!=i):
                similarity=Gaussian_similarity_f(data.loc[i],data.loc[j],sigma)
                nominator=nominator+c_i_j(labels,i,j)*similarity
                denominator=denominator+similarity
        tao=tao+nominator/denominator
    tao=tao/len(labels)
    return tao
```

Figure 1: Custom function for tao index

Table 1: Result of clustering

|  | SI | DB | NMI | $\tau$ |
|---|---|---|---|---|
| K-means | 0.3600 for K=3 | 0.8779 for K=3 | 0.0066 for K=5 | 0.9711 for K=2 |
| Spectral | 0.0253 for K=2 | 5.4585 for K=5 | 1.0 for K=3 | 0.9996 for K=2 |

We can see that according to $\tau$ index better clustering is still Spectral clustering but with K=2. I should mention that K=3 was very close with value 0.9994.

## 1.2  B)

**Task**: Compare the results with the ones indicated by Silhouette and Davies-Bouldin indices. Discuss the results. Is  index a better choice in this case?

**Answer**: As we can see $\tau$ index did better job than SI and DB in Spec-

Table 2: Result of clustering

|  | SI | DB | NMI | $\tau$ |
|---|---|---|---|---|
| K-means | 0.3600 for K=3 | 0.8779 for K=3 | 0.0066 for K=5 | 0.9711 for K=2 |
| Spectral | 0.0253 for K=2 | 5.4585 for K=5 | 1.0 for K=3 | 0.9996 for K=2 |

tral clustering but it failed for the K-means clustering. Overall i found that my implementation of $\tau$ index didn't perfom well since all of its values were quite high and quite close to each other.

*PS. I have tried hard and lost a lot of time on debugging for possible mistakes but couldn't find anything wrong in the code of $\tau$ index so if there is something wrong I would be very happy if I could get any feedback from the checking person. Thank you in advance!*