AML-Course---MID-SEM

Submitted by: - Hemant Yadav (PhD20017) (hemantya@iiit.ac.in)

Question 2:-

ARI: Adjusted Rand Index
AMI: Adjusted Mutual Information
NMI: Normalized Mutual Information

- **The code is present in the mid-sem.ipynb jupyter notebook. It is self explanatory. And the instructions to run it are in the README.md file.**
- **The average scores are averaged over 10 runs**
- **The dataset used as mentioned in the question.**
- **PCA is used to reduce the number of dimensions such that the total energy retained is not less than 95%. Initially we had 748 dimension which after the PCA operation reduced to 121**
- **1 and 2 K-means shown below are based on the kernel approach derived in question 1a. And 1b.**
- **3 is simply the SK-learn based K-means implementation.**

Observations:

- **The Gaussian kernel gives better results compared to the linear kernel.**
- **The Gaussian kernel t-SNE plot has less overlap between the different clusters compared to the Euclidean one.**
- **The reason is since the Gaussian kernel takes the basis to the infinite space and it's easy to make clusters in that space.**

The quantitative results are shown below:-

1. **Linear kernel K-means :-**

    a. Average score :
        i. ari:0.57
        ii. ami:0.61
        iii. nmi:0.61
    b. Best score :
        i. ari:0.64
        ii. ami:0.65
        iii. nmi:0.65

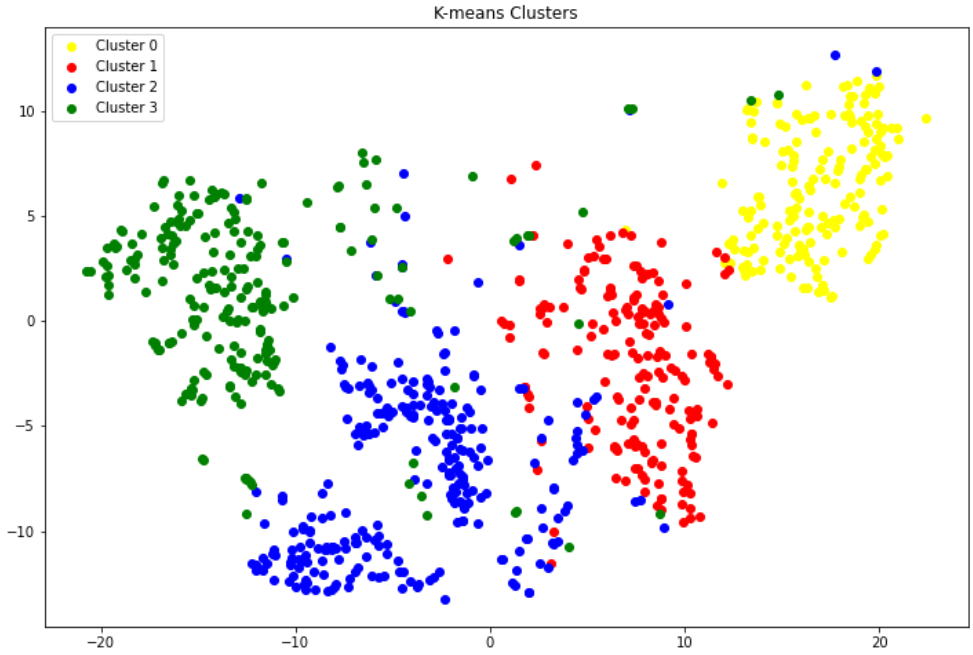

Figure 1. t-SNE plot of Linear Kernel implementation.

2. **Gaussian Kernel K-means:-** we ran the cross validation between (5 and 1e-10). The best value scale value was 1e-7

    a. Average score
        i. ari:0.59
        ii. ami:0.62
        iii. nmi:0.62
    b. Best score
        i. ari:0.73
        ii. ami:0.71
        iii. Nmi:0.71

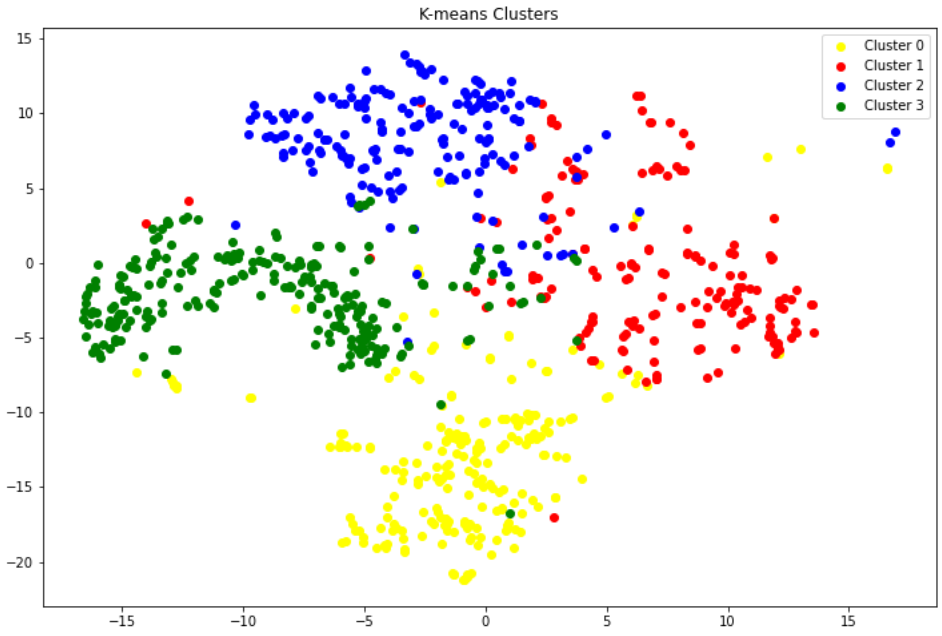

Figure 2. t-SNE plot of Gaussian Kernel implementation.

3. **Sk-learn usual K-means** (Euclidean distance calculation using points directly):-

    a. Average score:
        i. ari:0.47
        ii. ami:0.53
        iii. nmi:0.53
    b. Best score:
        i. ari:0.64
        ii. ami:0.65
        iii. Nmi:0.65

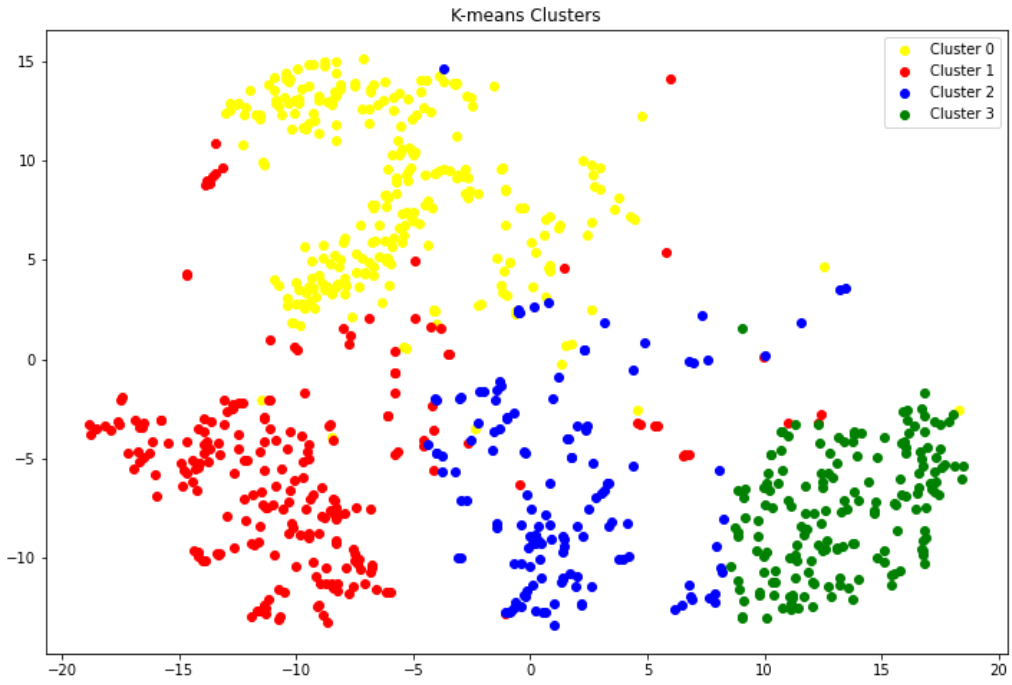

Figure 1. t-SNE plot of SK-learn implementation.