**Assignment 4: K-means**

**Results for 3 runs without data shuffling(Euclidean distance)**:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Run#** | **Cluster 0** | | | | **Cluster 1** | | | |
|  | Positive | Negative | Total | %Positive | Positive | Negative | Total | %Positive |
| Run 1 | 3 | 149 | 152 | 1.97 | 354 | 63 | 417 | 84.89 |
| Run 2 | 3 | 141 | 144 | 2.08 | 354 | 71 | 425 | 83.29 |
| Run 3 | 3 | 140 | 143 | 2.09 | 354 | 72 | 426 | 83.09 |

Here, we can observe that k-means clustering algorithm did not cluster the data perfectly. A perfect clustering would be when all sample in a cluster have same label. Here, we have a fraction of other class in the both the clusters.

**Results for 3 runs without data shuffling(Manhattan distance)**:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Run#** | **Cluster 0** | | | | **Cluster 1** | | | |
|  | Positive | Negative | Total | %Positive | Positive | Negative | Total | %Positive |
| Run 1 | 2 | 141 | 143 | 1.39 | 355 | 71 | 426 | 83.33 |
| Run 2 | 2 | 139 | 141 | 1.41 | 355 | 73 | 428 | 82.94 |
| Run 3 | 355 | 71 | 426 | 83.33 | 2 | 141 | 143 | 1.39 |

Here, same as above we can observe that k-means clustering algorithm did not cluster the data perfectly. Here, we have a fraction of other class in the both the clusters.

**Results for 3 runs with data shuffling(Manhattan distance)**:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Run#** | **Cluster 0** | | | | **Cluster 1** | | | |
|  | Positive | Negative | Total | %Positive | Positive | Negative | Total | %Positive |
| Run 1 | 3 | 143 | 146 | 2.05 | 354 | 69 | 423 | 83.68 |
| Run 2 | 3 | 149 | 152 | 1.97 | 354 | 63 | 417 | 84.89 |
| Run 3 | 3 | 140 | 143 | 2.09 | 354 | 72 | 426 | 83.09 |

As from comparing with above Manhattan results, we can see that we do not see any significant improvement or decrement in drop of accuracy. Therefore, shuffling is not a mandatory step for k-means.