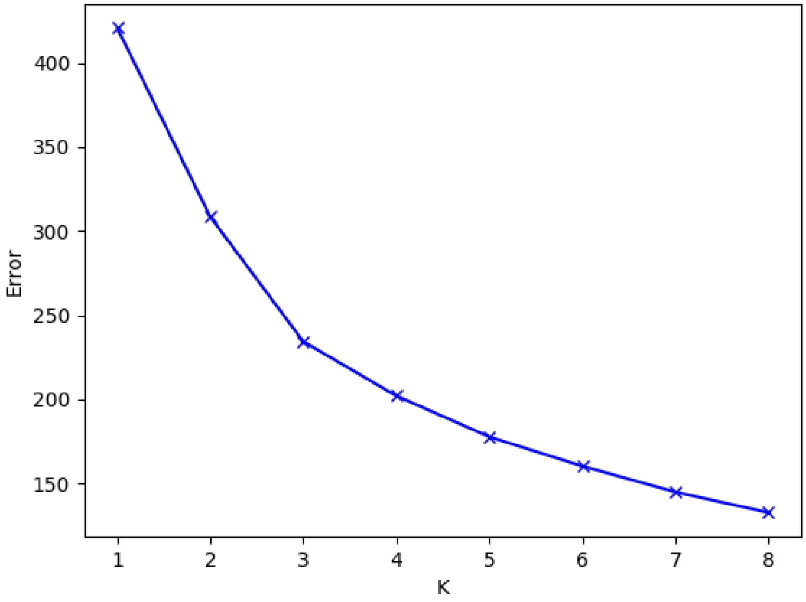
Home work 4

Part -1

1. Scipy.cluster.vq.kmeans for creating clusters, and vq.vq to create the code book for vector quantization
2. Used sklearn.cluster.kmeans for finding kmeans
3. Scipy.cluster.hierarchy.dendograms to create the dendograms

Results

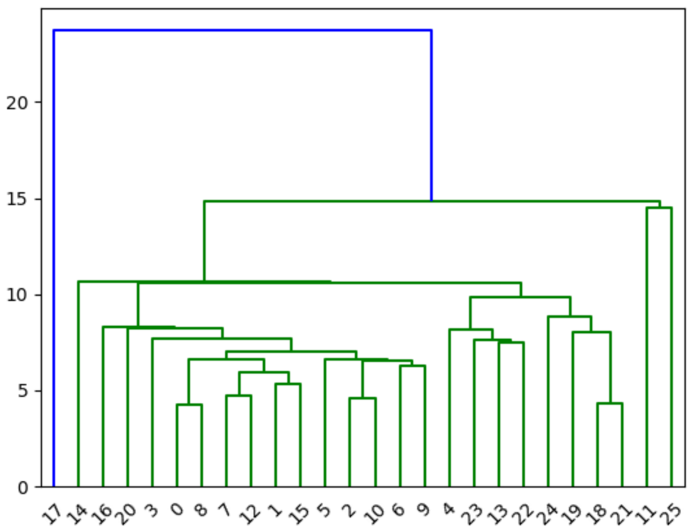
Kmeans error plot



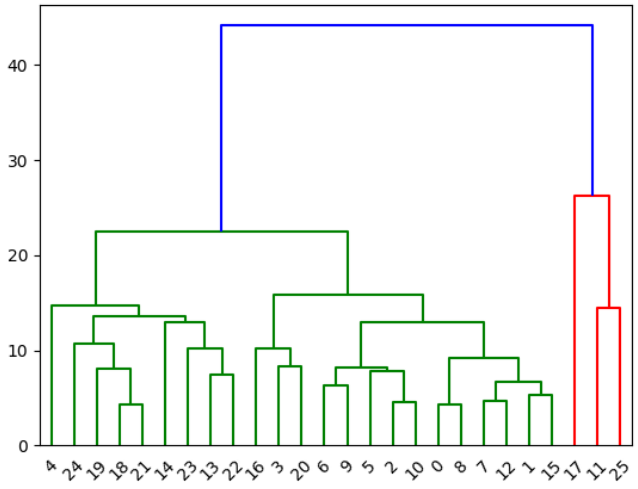
From above error plot we see that after K=3 is where the elbow starts and seems like a good value for K. From this point onwards the shift in error with increasing K reduces and eventually it tapers to constant value

All graphs below are against class clusters

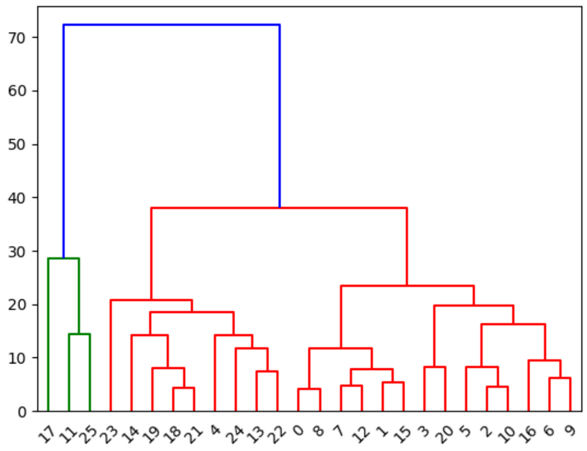
Single linkage



Average linkage



Complete linkage



Part -2

Software used

4. Scipy.cluster.vq.kmeans for creating clusters, and vq.vq to create the code book for vector quantization
5. sklearn.ensemble.RandomForestClassifier - for classification
6. Numpy and pandas for data manipulation
7. For each file I created histograms, then split it into training and test data
8. Using Random Forest to run the classification and finally get results
9. I ran several versions of runs with different splits and K's for kmeans. For all of them my accuracy was between 70-75%.
10. Here is a run down of the results

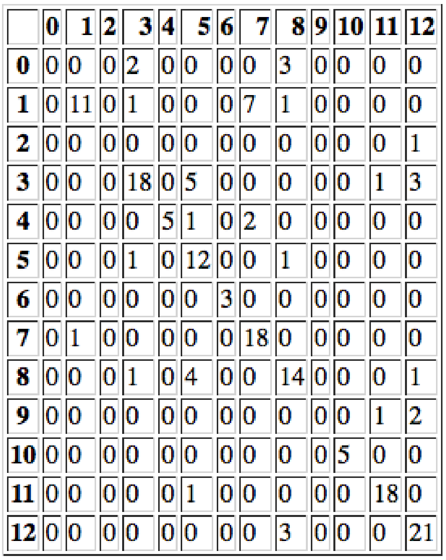
    Assumption :- The figure belows list class labels as int. They represnts actual class labels as follows

    - 0-Liedown_bed
    - 1-Walk
    - 2 -Eat_soup
    - 3-Getup_bed
    - 4-Descend_stairs
    - 5-UseTephone
    - 6-Standup_cahair
    - 7-Brush_teeth
    - 8-Climb_stairs
    - 9-Sitdown_chair
    - 10-Eat_meat
    - 11-Comb_hair

- 12-Drink_glass
- 13-Pour_water

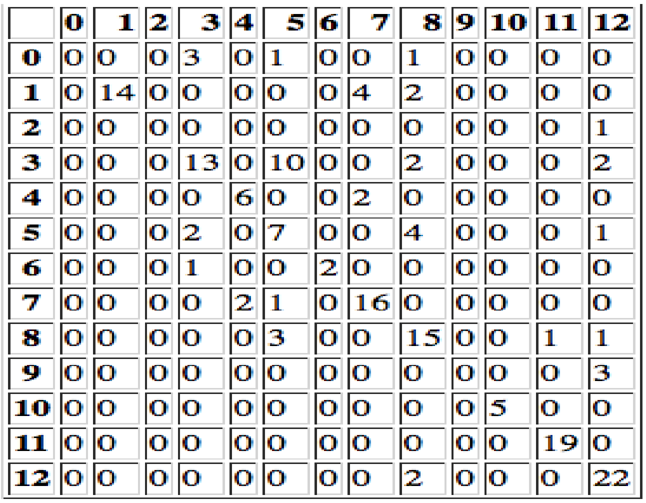
Results:-

1. K=480 Split 32
    a. Accuracy 74%
    b. Confusion matrix



| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 |
| **1** | 0 | 11 | 0 | 1 | 0 | 0 | 0 | 7 | 1 | 0 | 0 | 0 | 0 |
| **2** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **3** | 0 | 0 | 0 | 18 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 1 | 3 |
| **4** | 0 | 0 | 0 | 0 | 5 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| **5** | 0 | 0 | 0 | 1 | 0 | 12 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| **6** | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| **7** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 18 | 0 | 0 | 0 | 0 | 0 |
| **8** | 0 | 0 | 0 | 1 | 0 | 4 | 0 | 0 | 14 | 0 | 0 | 0 | 1 |
| **9** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| **10** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| **11** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 18 | 0 |
| **12** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 21 |

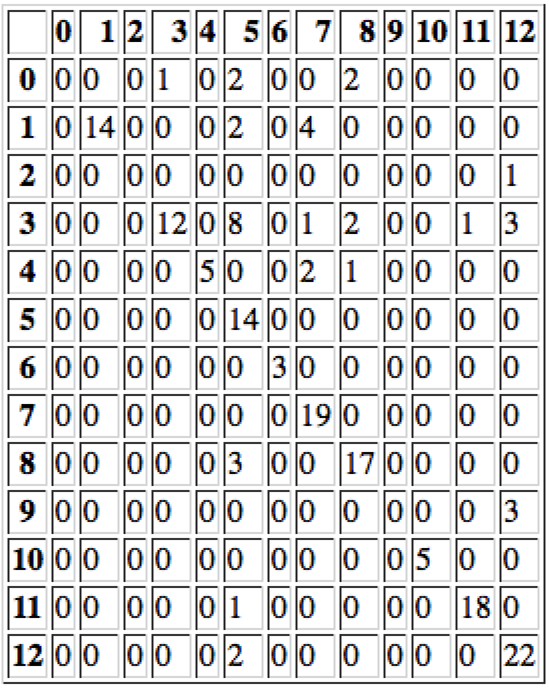
2. K=500 Split 32
    a. Accuracy 70%
    b. Confusion matrix



| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| **1** | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 4 | 2 | 0 | 0 | 0 | 0 |
| **2** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **3** | 0 | 0 | 0 | 13 | 0 | 10 | 0 | 0 | 2 | 0 | 0 | 0 | 2 |
| **4** | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| **5** | 0 | 0 | 0 | 2 | 0 | 7 | 0 | 0 | 4 | 0 | 0 | 0 | 1 |
| **6** | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| **7** | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 16 | 0 | 0 | 0 | 0 | 0 |
| **8** | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 15 | 0 | 0 | 1 | 1 |
| **9** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| **10** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| **11** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 0 |
| **12** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 22 |

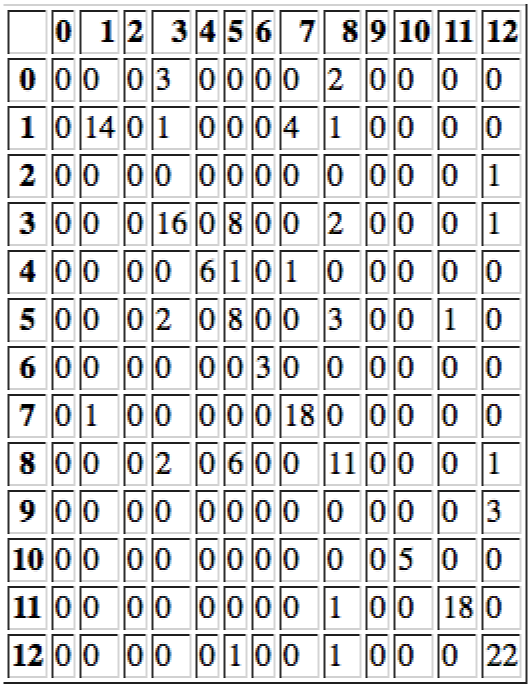
3. K=480 Split 25

a. Accuracy 76%
b. Confusion matrix



| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| **1** | 0 | 14 | 0 | 0 | 0 | 2 | 0 | 4 | 0 | 0 | 0 | 0 | 0 |
| **2** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **3** | 0 | 0 | 0 | 12 | 0 | 8 | 0 | 1 | 2 | 0 | 0 | 1 | 3 |
| **4** | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 |
| **5** | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **6** | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| **7** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 |
| **8** | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 17 | 0 | 0 | 0 | 0 |
| **9** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| **10** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| **11** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 18 | 0 |
| **12** | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 22 |

4. K=300 Split 20
   a. Accuracy 72%
   b. Confusion matrix



| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| **1** | 0 | 14 | 0 | 1 | 0 | 0 | 0 | 4 | 1 | 0 | 0 | 0 | 0 |
| **2** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **3** | 0 | 0 | 0 | 16 | 0 | 8 | 0 | 0 | 2 | 0 | 0 | 0 | 1 |
| **4** | 0 | 0 | 0 | 0 | 6 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| **5** | 0 | 0 | 0 | 2 | 0 | 8 | 0 | 0 | 3 | 0 | 0 | 1 | 0 |
| **6** | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| **7** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 18 | 0 | 0 | 0 | 0 | 0 |
| **8** | 0 | 0 | 0 | 2 | 0 | 6 | 0 | 0 | 11 | 0 | 0 | 0 | 1 |
| **9** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| **10** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| **11** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 18 | 0 |
| **12** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 22 |

We see that results dont change much with changing splits and K’s and accuracy remains about the same.