ASSIGNMENT III

MUHAMMAD KHATTAB

34-14154

ASSOC. PROF. SEIF ELDAWLATLY

CONTENTS

1 INTRODUCTION

Provided with total 182 images, the objective is to examine the performance of different clustering algorithms available in Weka on the data. Each row in the file represents one image (total 182 images) and each column represents the brightness of one pixel (total 144 pixels). The last column represents the character that is present in each image. Ideally, the clustering algorithm should cluster the 7 images corresponding to each character in a separate cluster. The clustering algorithms checked are:

- K-means clustering,
- hierarchical clustering with single linkage,
- hierarchical clustering with average linkage,
- and hierarchical clustering with complete linkage.

2 K-MEANS CLUSTERING

Table 1: Results of clustering into 26 clusters

Cluster	No. Instances	Percentage
0	13	7%
1	6	3%
2	12	7%
3	3	2%
4	5	3%
5	3	2%
6	7	4%
7	2	1%
8	13	7%
9	9	5%
10	2	1%
11	3	2%
12	1	12%
13	1	1%
14	9	5%
15	4	2%
16	16	9%
17	1	1%
18	2	1%
19	3	2%
20	7	5%
21	2	1%
22	11	6%
23	11	6%
24	2	1%
25	14	8%

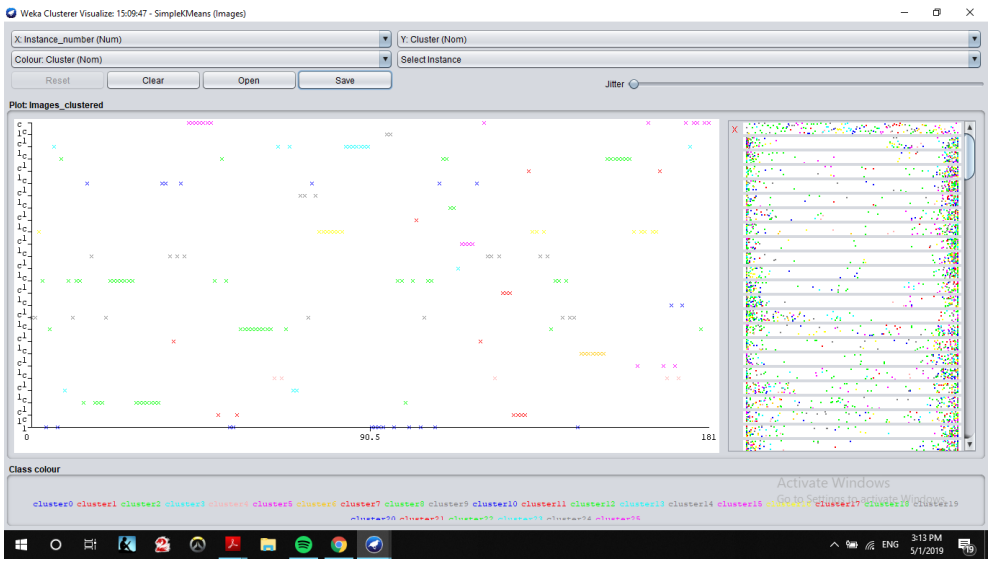

Figure 1: K-means clustering

3 HIERARCHICAL CLUSTERING WITH SINGLE LINKAGE

Table 2: Results of clustering into 26 clusters

Cluster	No. Instances	Percentage
0	5	3%
1	2	1%
2	136	75%
3	1	1%
4	1	1%
5	1	1%
6	1	1%
7	1	1%
8	1	1%
9	1	1%
10	1	1%
11	7	4%
12	1	1%
13	1	1%
14	1	1%
15	1	1%
16	1	1%
17	6	3%
18	1	1%
19	1	1%
20	1	1%
21	1	1%
22	1	1%
23	5	3%
24	1	1%
25	1	1%

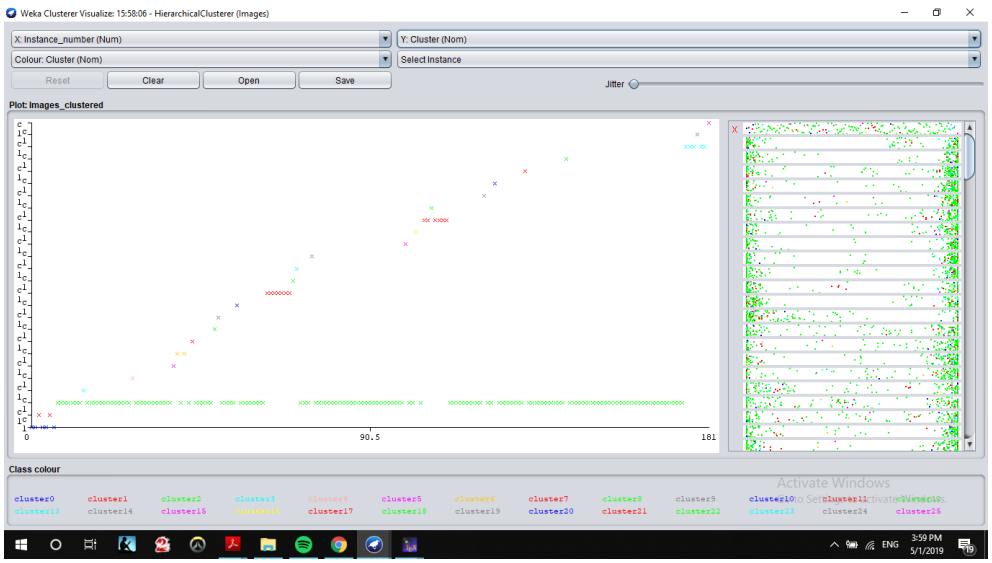

Figure 2: Hierarchical Clustering with Single Linkage

4 HIERARCHICAL CLUSTERING WITH AVERAGE LINKAGE

Table 3: Results of clustering into 26 clusters

Cluster	No. Instances	Percentage
0	7	4%
1	9	5%
2	7	4%
3	7	4%
4	9	5%
5	9	5%
6	1	1%
7	2	1%
8	7	4%
9	19	10%
10	9	5%
11	1	1%
12	21	12%
13	7	4%
14	6	3%
15	8	4%
16	1	1%
17	7	4%
18	7	4%
19	2	1%
20	6	3%
21	1	1%
22	1	1%
23	14	8%
24	7	4%
25	7	4%

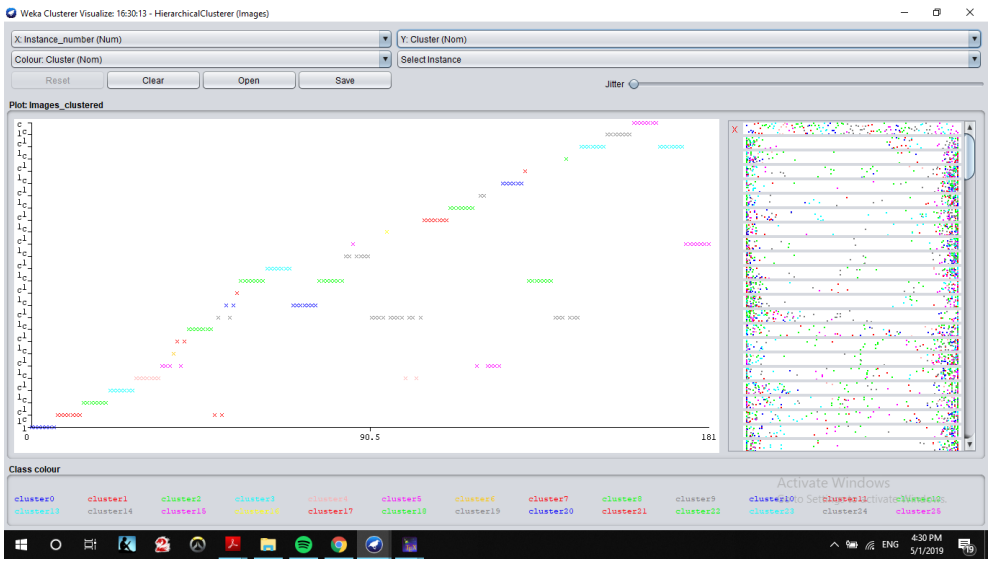

Figure 3: Hierarchical Clustering with Average Linkage

5 HIERARCHICAL CLUSTERING WITH COMPLETE LINKAGE

Table 4: Results of clustering into 26 clusters

Cluster	No. Instances	Percentage
0	7	4%
1	7	4%
2	5	3%
3	7	4%
4	7	4%
5	9	5%
6	5	3%
7	6	3%
8	7	4%
9	5	3%
10	4	2%
11	5	3%
12	9	5%
13	7	4%
14	7	4%
15	12	7%
16	6	3%
17	8	4%
18	13	7%
19	7	4%
20	8	4%
21	2	1%
22	6	3%
23	2	1%
24	14	8%
25	7	4%

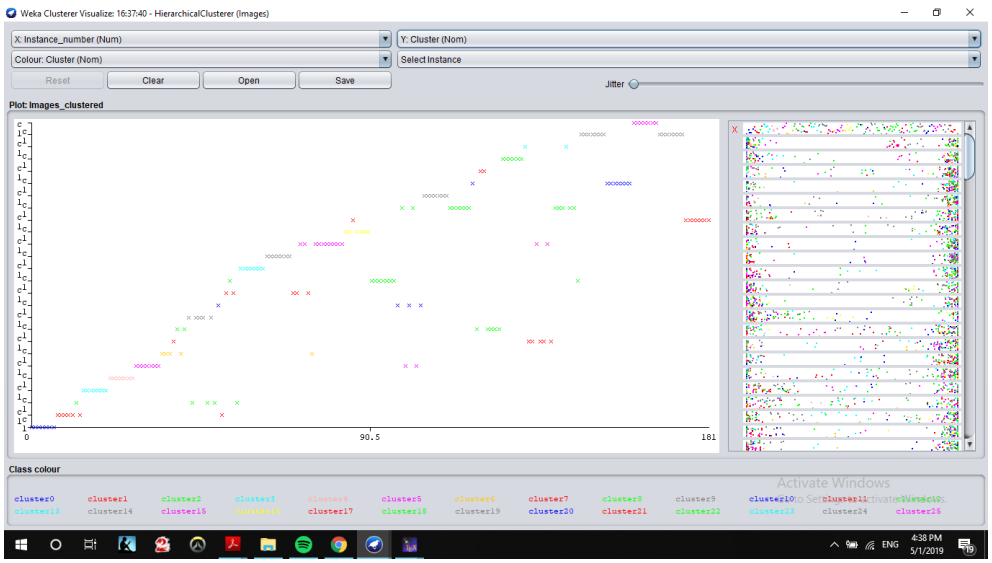

Figure 4: Hierarchical Clustering with Complete Linkage

6 CONCLUSION

Hierarchical clustering with single linkage performs worst of the four algorithms. As 136 (75%) of the input gets assigned to a single cluster which should be only 7, as per table 2 and figure 2.

K-means clustering performed better than the hierarchical clustering with single linkage. However, a number of cluster contain only 1 or 2 instances, as per table 1 and figure 1.

Hierarchical clustering with average linkage and complete linkage provide similar clusters, to some extent. However, the number of misclassified instances is a bit less in complete linkage, by comparing table 3, figure 3, table 4, and figure 4. Meaning that hierarchical clustering with complete linkage provides the best results.