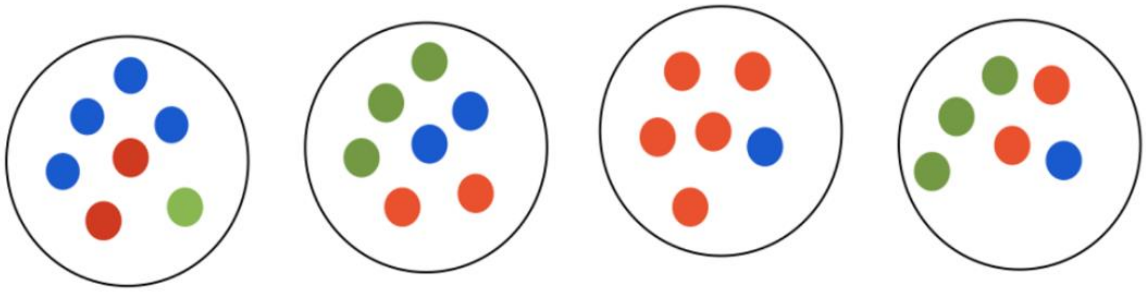
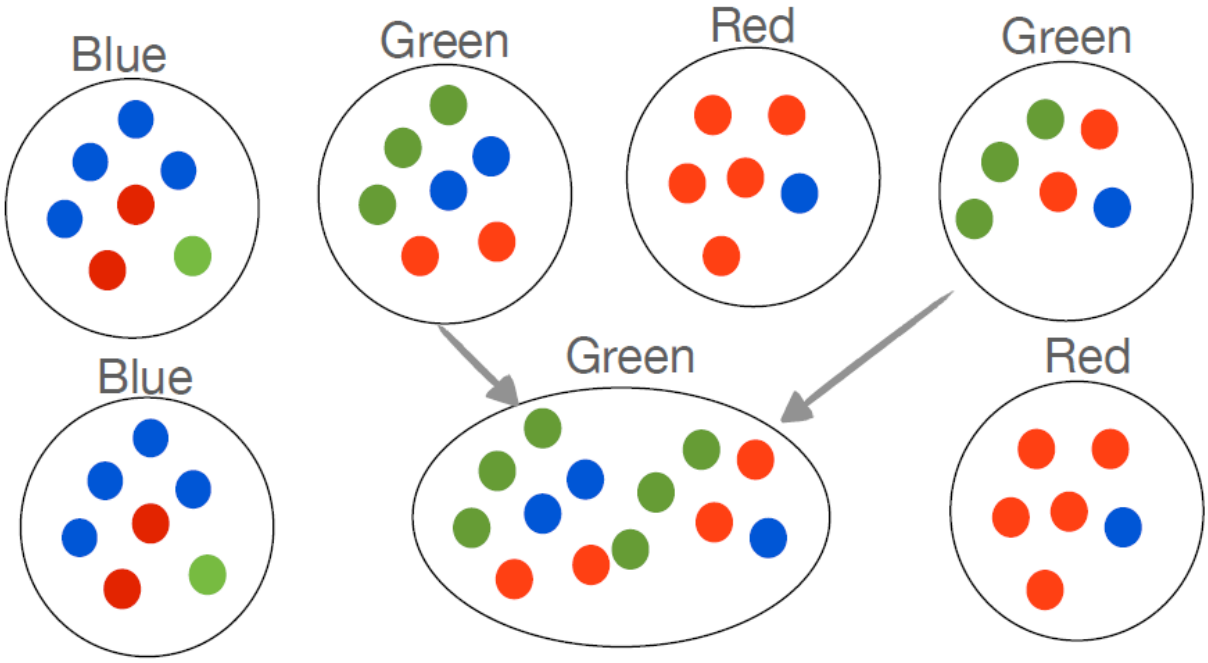
Name: Afolabi, Akinola Olawale
StudentID: 201770868

Report on Questions 5 and 6 of CourseWork 2

From figure 1,



Assign each cluster the label that appears most in that cluster and merge clusters with the same label.



From this clusters, the confusion matrix is given as:

		Actual		
	Labels	Blue	Green	Red
Predicted	Blue	4	1	2
	Green	3	6	4
	Red	1	0	5

$$Precision\ for\ Blue\ Label = \frac{no\ of\ objects\ correctly\ classified\ blue}{no\ of\ objects\ classified\ blue}$$

$$Precision\ for\ Blue\ Label = \frac{4}{7}$$

$$Precision\ for\ Blue\ Label = 0.571429$$

$$Precision\ for\ Green\ Label = \frac{no\ of\ objects\ correctly\ classified\ green}{no\ of\ objects\ classified\ green}$$

$$Precision\ for\ Green\ Label = \frac{6}{13}$$

$$Precision\ for\ Green\ Label = 0.461538$$

$$Precision\ for\ Red\ Label = \frac{no\ of\ objects\ correctly\ classified\ red}{no\ of\ objects\ classified\ red}$$

$$Precision\ for\ Red\ Label = \frac{5}{6}$$

$$Precision\ for\ Red\ Label = 0.833333$$

$$Macro - averaged\ Precision = \frac{1}{n}\sum_{i=1}^{n} P_i$$

Where Pi – Precision for each label class and n – number of label class

$$Macro - averaged\ Precision = \frac{0.571429 + 0.833333 + 0.461538}{3}$$

$$Macro - averaged\ Precision = 0.622$$

$$Recall\ for\ Blue\ Label = \frac{no\ of\ objects\ correctly\ classified\ blue}{no\ of\ objects\ that\ belongs\ to\ blue\ cluster}$$

$$Recall\ for\ Blue\ Label = \frac{4}{8}$$

$$Recall\ for\ Blue\ Label = 0.5$$

$$Recall\ for\ Green\ Label = \frac{no\ of\ objects\ correctly\ classified\ green}{no\ of\ objects\ that\ belongs\ to\ green\ cluster}$$

$$Recall\ for\ Green\ Label = \frac{6}{7}$$

$$Recall\ for\ Green\ Label = 0.857143$$

$$Recall\ for\ Red\ Label = \frac{no\ of\ objects\ correctly\ classified\ red}{no\ of\ objects\ that\ belongs\ to\ red\ cluster}$$

$$Recall\ for\ Red\ Label = \frac{5}{11}$$

$$Recall\ for\ Red\ Label = 0.454545$$

$$Macro - averaged\ Recall = \frac{1}{n}\sum_{i=1}^{n} R_i$$

Where Ri – Recall for each label class and n – number of label class

$$Macro - averaged\ Recall = \frac{0.5 + 0.857143 + 0.454545}{3}$$

$$Macro - averaged\ Recall = 0.604$$

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$F - score\ for\ Blue\ Label = \frac{2 \times 0.5 \times 0.571429}{0.5 + 0.571429}$$

$$F - score\ for\ Blue\ Label = 0.533334$$

$$F - score\ for\ Green\ Label = \frac{2 \times 0.857143 \times 0.461538}{0.857143 + 0.461538}$$

$$F - score\ for\ Blue\ Label = 0.60$$

$$F - score\ for\ Red\ Label = \frac{2 \times 0.454545 \times 0.833333}{0.454545 + 0.833333}$$

$$F - score\ for\ Red\ Label = 0.588235$$

$$Macro\ F - score = \frac{1}{n}\sum_{i=1}^{n} F - score_i$$

Where F-Score$_i$ – F-Score for each label class and n – number of label class

$$Macro\ F - score = \frac{0.588235 + 0.6 + 0.533334}{3}$$

$$Macro\ F - score = 0.574$$

B-cubed Precision, Recall and F-score is given as

$$Precision = \frac{no\ of\ items\ in\ C(x) with\ A(x)}{no\ of\ items\ in\ C(x)}$$

$$Recall = \frac{no\ of\ items\ in\ C(x) with\ A(x)}{total\ no\ of\ items\ with\ A(x)}$$

Where C(x) – ID of the cluster that x belongs to and A(x) – label of x

The precision for blue cluster with 4 data points of label blue is given by:

$$Precision\ for\ Blue\ Label = \frac{4}{7} \times 4$$

$$Precision\ for\ Blue\ Label = 2.2857$$

The precision for blue cluster with 1data point of label green is given by:

$$Precision\ for\ Green\ Label = \frac{1}{7}$$

$$Precision\ for\ Green\ Label = 0.1429$$

The precision for blue cluster with 2data points of label red is given by:

$$Precision\ for\ Red\ Label = \frac{2}{7} \times 2$$

$$Precision\ for\ Red\ Label = 0.5714$$

The precision for green cluster with 3 data points of label blue is given by:

$$Precision\ for\ Blue\ Label = \frac{3}{13} \times 3$$

$$Precision\ for\ Blue\ Label = 0.6923$$

The precision for green cluster with 6 data points of label green is given by:

$$Precision\ for\ Green\ Label = \frac{6}{13} \times 6$$

$$Precision\ for\ Green\ Label = 2.7692$$

The precision for green cluster with 4 data points of label red is given by:

$$Precision\ for\ red\ Label = \frac{4}{13} \times 4$$

$$Precision\ for\ red\ Label = 1.2308$$

The precision for red cluster with 1 data point of label blue is given by:

$$Precision\ for\ blue\ Label = \frac{1}{6} \times 1$$

$$Precision\ for\ blue\ Label = 0.1667$$

The precision for red cluster with 5 data points of red label is given by:

$$Precision\ for\ red\ Label = \frac{5}{6} \times 5$$

$$Precision\ for\ red\ Label = 4.1667$$

The average Bcubed precision is given by:

$$Precision = \frac{2.2857 + 0.1429 + 0.5714 + 0.6923 + 2.7692 + 1.2308 + 0.1667 + 4.1667}{26}$$

$$Average\ Bcubed\ Precision = 0.4625$$

The recall for blue cluster with 4 data points of label blue is given by:

$$Recall\ for\ Blue\ Label = \frac{4}{8} \times 4$$

$$Recall\ for\ Blue\ Label = 2.0000$$

The recall for blue cluster with 1data point of label green is given by:

$$Recall\ for\ Green\ Label = \frac{1}{7}$$

$$Recall\ for\ Green\ Label = 0.1429$$

The recall for blue cluster with 2data points of label red is given by:

$$Recall\ for\ Red\ Label = \frac{2}{11} \times 2$$

$$Recall\ for\ Red\ Label = 0.3636$$

The recall for green cluster with 3 data points of label blue is given by:

$$Recall\ for\ Blue\ Label = \frac{3}{8} \times 3$$

$$Recall\ for\ Blue\ Label = 1.125$$

The recall for green cluster with 6 data points of label green is given by:

$$Recall\ for\ Green\ Label = \frac{6}{7} \times 6$$

$$Recall\ for\ Green\ Label = 5.1429$$

The recall for green cluster with 4 data points of label red is given by:

$$Recall\ for\ red\ Label = \frac{4}{11} \times 4$$

$$Recall\ for\ red\ Label = 1.4545$$

The recall for red cluster with 1 data point of label blue is given by:

$$Recall\ for\ blue\ Label = \frac{1}{8} \times 1$$

$$Recall\ for\ blue\ Label = 0.1250$$

The recall for red cluster with 5 data points of red label is given by:

$$Recall\ for\ red\ Label = \frac{5}{11} \times 5$$

$$Recall\ for\ red\ Label = 2.2727$$

The Average Bcubed Recall is given by:

$$Recall = \frac{2 + 0.1429 + 0.3636 + 1.125 + 5.1429 + 1.4545 + 0.1250 + 2.2727}{26}$$

$$Average\ Bcubed\ Recall = 0.4856$$

The Average Bcubed F-score is given by:

$$F - score = \frac{2 \times\ recall \times precision}{recall + precision}$$

$$Average\ Bcubed\ F - score = \frac{1}{n}\sum_{i=1}^{n} F - score_i$$

The F-score for Cluster 1 with 4 data points of label blue is given by:

$$F - score\ for\ Blue\ Label = \frac{2\ \times 2.2857\ \ \times 2.000}{2.2857 + 2.000}$$

$$F - score\ for\ Blue\ Label = 2.1333$$

The F-score for Cluster 1 with 1 data point of label green is given by:

$$F - score\ for\ Green\ Label = \frac{2\ \times 0.1429\ \ \times 0.1429}{0.1429 + 0.1429}$$

$$F - score\ for\ Green\ Label = 0.1429$$

The F-score for Cluster 1 with 2 data points of label red is given by:

$$F - score\ for\ Red\ Label = \frac{2\ \times 0.5714\ \ \times 0.3636}{0.5714 + 0.3636}$$

$$F - score\ for\ Red\ Label = 0.4444$$

The F-score for cluster 2 with 3 data points of label blue is given by:

$$F - score\ for\ Blue\ Label = \frac{2\ \times 1.125\ \ \times 0.6923}{1.125 + 0.6923}$$

$$F - score\ for\ Blue\ Label = 0.8571$$

The F-score for cluster 2 with 6 data points of label green is given by:

$$F - score\ for\ Green\ Label = \frac{2\ \times 2.7692\ \ \times 5.1429}{2.7692 + 5.1429}$$

$$F - score\ for\ Green\ Label = 3.600$$

The F-score for cluster 2 with 4 data points of label red is given by:

$$F - score\ for\ Red\ Label = \frac{2\ \times 1.2308\ \ \times 1.4545}{1.2308 + 0.4545}$$

$$F - score\ for\ Red\ Label = 1.333$$

The F-score for cluster 3 with 1 data point of label blue is given by:

$$F - score\ for\ Blue\ Label = \frac{2\ \times 0.125\ \ \times 0.1667}{0.125 + 0.1667}$$

$$F - score\ for\ Blue\ Label = \ 0.1429$$

The F-score for cluster 3 with 5 data points of red label is given by:

$$F - score\ for\ Red\ Label = \frac{2\ \times 2.2727\ \ \times 4.1667}{2.2727 + 4.1667}$$

$$F - score\ for\ Red\ Label = 2.941$$

The Average Bcubed F-score is given by:

$$F - score = \frac{2.1333 + 0.1429 + 0.4444 + 0.8571 + 3.6 + 1.333 + 0.1429 + 2.9412}{26}$$

$$Average\ Bcubed\ F - score = 0.446$$

The Bcubed and Macro-Averaged Precision, Recall and F-score is summarized in the table below:

Metrics	Macro-Average	B-Cubed
Precision	0.622	0.463
Recall	0.604	0.486
F-score	0.574	0.446