Assignment 2
Marzhan Nuraly

Task 5

The clustering outcomes shown in Figure 1. Then each cluster is assigned to the label that appears most in the cluster. After that clusters with the same label merged as shown in Figure 2.

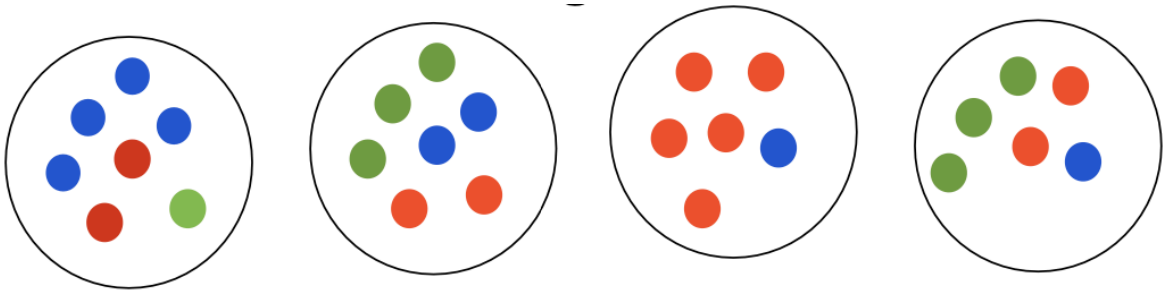

Figure 1

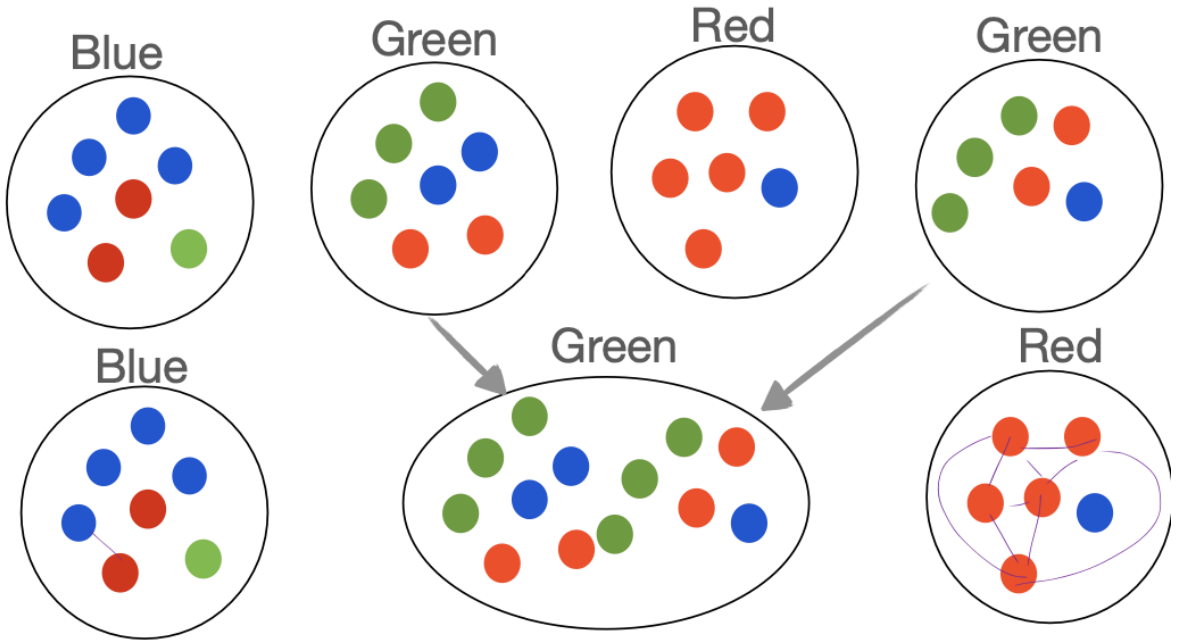

Figure 2

The confusion matrix for whole classes is shown in Table 1 and Table 2.

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Table 1

Assignment 2
Marzhan Nuraly

		Actual	
		Positive	Negative
Predicted	Positive	41	73
	Negative	63	121

Table 2

Evaluation measures for multiple classes:

Precision for class A: $\frac{no.\ objects\ correctly\ classifies\ A}{no.\ objects\ classified\ A}$

Recall for class A: $\frac{no.\ objects\ correctly\ classifies\ A}{no.\ objects\ that\ belongs\ to\ class\ A}$

Precision $\Rightarrow Pr_{blue} = \frac{4}{7}, Pr_{green} = \frac{6}{13}, Pr_{red} = \frac{5}{6}$, **Recall** $\Rightarrow Rc_{blue} = \frac{4}{8}, Rc_{green} = \frac{6}{7}, Rc_{red} = \frac{5}{11}$

F-score $\Rightarrow Fc_{blue} = \frac{2*Pr_{blue}*Rc_{blue}}{Pr_{blue}+Rc_{blue}} = 0.533, Fc_{green} = \frac{2*Pr_{green}*Rc_{green}}{Pr_{green}+Rc_{green}} = 0.600$

$$Fc_{red} = \frac{2*Pr_{red}*Rc_{red}}{Pr_{red}+Rc_{red}} = 0.587$$

The formula for finding macro-average precision:

$$\text{Macro-average Precision} = \frac{Pr_{blue} + Pr_{green} + Pr_{red}}{3}$$

The formula for finding macro-average recall:

$$\text{Macro-average Recall} = \frac{Rc_{blue} + Rc_{green} + Rc_{red}}{3}$$

The formula for finding macro-average F-score:

$$\text{Macro-average F-score} = \frac{1}{c}\sum_{i=1}^{c} F_{score_i}$$

The Results:

Macro-average Precision for the Figure 2:

$$Pr = \frac{1}{3} * (\frac{4}{7} + \frac{6}{13} + \frac{5}{6}) = 0.621$$

Macro-average Recall:

$$Rc = \frac{1}{3} * (\frac{4}{8} + \frac{6}{7} + \frac{5}{11}) = 0.603$$

Macro-average F-score:

$$Fc = \frac{0.533+0.600+0.587}{3} = 0.573$$

Micro-average Precision:

$$Precision = \frac{TP}{TP + FP} => Pr = \frac{41}{41 + 73} = 0.359$$

Micro-average Recall:

$$Recall = \frac{TP}{TP + FN} => Rc = \frac{41}{41 + 63} = 0.3942$$

F-score:

$$F_{score} = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} => F_{score} = \frac{2 * 0.3942 * 0.359}{0.3942 + 0.359} = 0.3757$$

Task 6

The formula to calculate B-CUBED measurements is given below:

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

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

C(x) - the ID of the cluster that x belongs to, *A(x)* - label of x

$$Precision = \frac{1}{N} \sum_{p\ in\ dataset} Precision\ (p)$$

$$Recall = \frac{1}{N} \sum_{p\ in\ dataset} Recall\ (p)$$

$$F_{score} = \frac{1}{N} \sum_{p\ in\ dataset} F_{score}\ (p)$$

Initially the precision and recall is computed for each item:

Precision for Blue points:

$$4 * \frac{4}{7} = 2.285, 3 * \frac{3}{13} = 0.692, 1 * \frac{1}{6} = 0.166$$

Precision for Green points:

$$1 * \frac{1}{7} = 0.142, 6 * \frac{6}{13} = 2.769$$

Precision for Red points:

$$2 * \frac{2}{7} = 0.571, 4 * \frac{4}{13} = 1.230, 5 * \frac{5}{7} = 4.166$$

The results:

$$B - CUBED\ Precision = \frac{2.285+0.692+0.166+0.142+2.769+0.571+1.230+4.166}{26} = 0.462$$

Recall for Blue points:

$$4 * \frac{4}{8} = 2, 3 * \frac{3}{8} = 1.125, 1 * \frac{1}{8} = 0.125$$

Recall for Green points:

$$1 * \frac{1}{7} = 0.142, 6 * \frac{6}{7} = 5.142$$

Recall for Red points:

$$2 * \frac{2}{11} = 0.363, 4 * \frac{4}{11} = 1.454, 5 * \frac{5}{11} = 2.272$$

The results:

$$B - CUBED\ Recall = \frac{2+1.125+0.125+0.142+5.142+0.363+1.454+2.272}{26} = 0.485$$

The results:

$$B - CUBED\ Fscore = \frac{2 * Precision * Recall}{Precision+Recall} = \frac{2 * 0.462 * 0.485}{0.462 + 0.485} = 0.473$$