

## Problem set 6

### Clustering

#### Exercise 1

Compute the confusion matrix, macro-averaged Precision, Recall, and F-score for the clustering shown below.

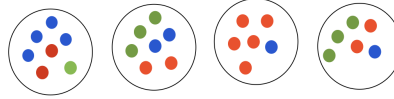


Figure 1: Caption

**Solution** First cluster will be labelled as blue. Second cluster will be labelled as green. Third cluster will be labelled as red. Fourth cluster will be labelled as green.

Hence Second and fourth cluster will be combined.

Confusion matrix is given below.

$$\begin{pmatrix} & \text{blue} & \text{red} & \text{green} \\ \text{blue} & 4 & 1 & 3 \\ \text{red} & 2 & 5 & 4 \\ \text{green} & 1 & 0 & 6 \end{pmatrix}$$

Now for blue labelled cluster precision is  $\frac{4}{7}$ . For green labelled cluster, the precision is  $\frac{6}{13}$ . For red labelled cluster the precision is  $\frac{5}{6}$ . Hence macro averaged precision is  $(0.571 + 0.461 + 0.833)/3 = 0.621$

Now for blue class label recall is  $\frac{4}{8}$ . For green label, the recall is  $\frac{6}{7}$ . For red label, the recall is  $\frac{5}{11}$ . Hence macro averaged recall is  $(0.5 + 0.857 + 0.454)/3 = 0.603$

Macro averaged FScore is  $(0.533 + 0.599 + 0.587)/3 = 0.573$

#### Exercise 2

For the same clusters as in Exercise 1, compute B-CUBED Precision, Recall, and F-score.

$$\text{BCubed Precision} = \frac{1}{26} \sum_{i=1}^{26} (4 * \frac{4}{7} + 2 * \frac{2}{7} + \frac{1}{7} + 6 * \frac{6}{13} + 3 * \frac{3}{13} + 4 * \frac{4}{13} + 5 * \frac{5}{6} + \frac{1}{6}) = 0.462$$

The recall and FScore can be computed in a similar way.