Normalized Mutual Information

Estimating Clustering Quality

Normalized Mutual Information

- Normalized Mutual Information:

$$NMI(Y,C) = \frac{2 \times I(Y;C)}{[H(Y) + H(C)]}$$

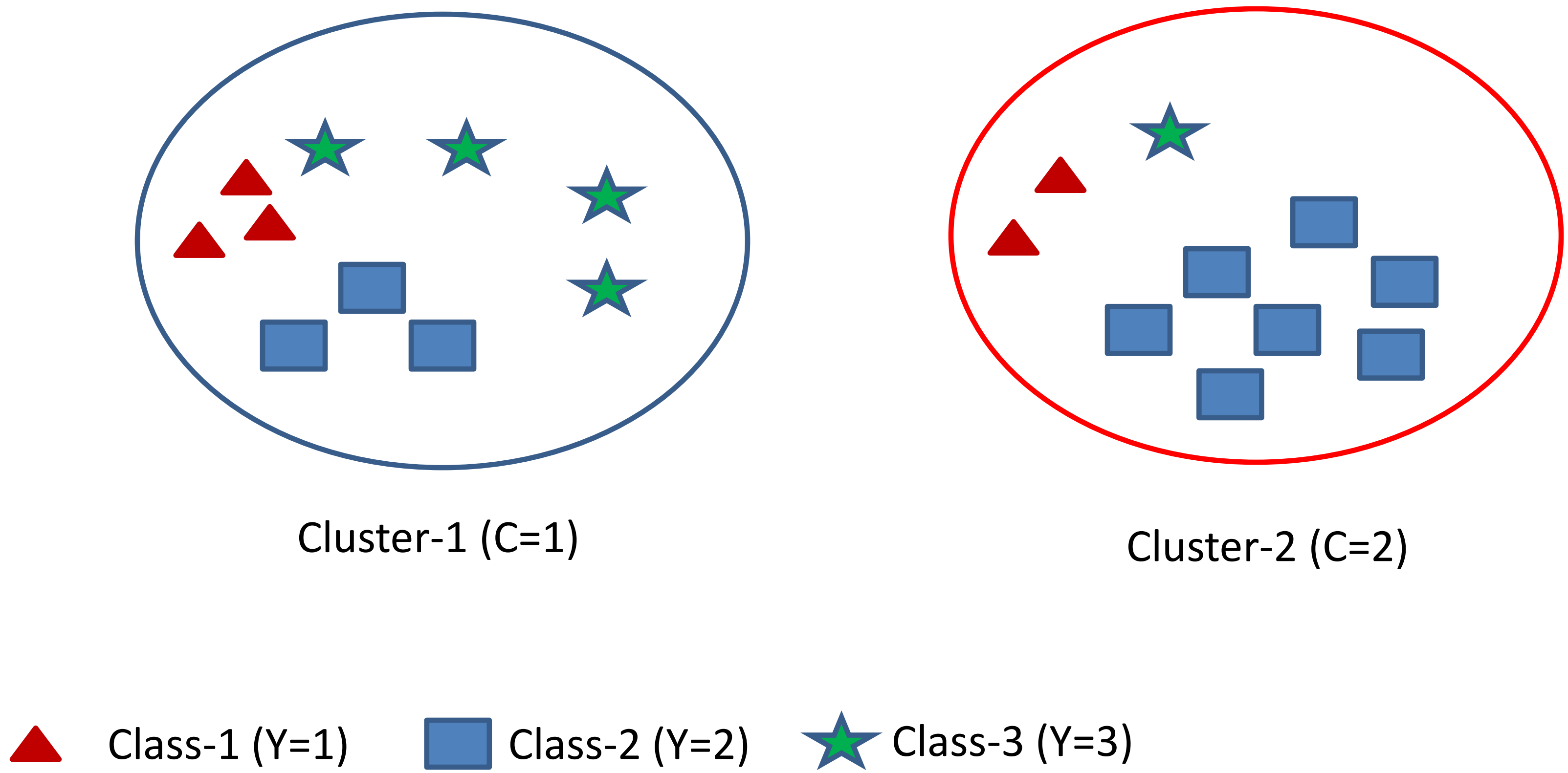
where,

1) Y = class labels
2) C = cluster labels
3) H(.) = Entropy
4) I(Y;C) = Mutual Information b/w Y and C

Note: All logs are base-2.

Calculating NMI for Clustering

- Assume m=3 classes and k=2 clusters



H(Y) = Entropy of Class Labels

- P(Y=1) = 5/20 = ¼
- P(Y=2) = 5/20 = ¼
- P(Y=3) = 10/20 = ½
- H(Y) = $-\frac{1}{4}\log\left(\frac{1}{4}\right)-\frac{1}{4}\log\left(\frac{1}{4}\right)-\frac{1}{2}\log\left(\frac{1}{2}\right)=1.5$

This is calculated for the entire dataset and can be calculated prior to clustering, as it will not change depending on the clustering output.

H(C) = Entropy of Cluster Labels

- P(C=1) = 10/20 = 1/2
- P(C=2) = 10/20 = ½
- H(Y) $= -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$

This will be calculated every time the clustering changes. You can see from the figure that the clusters are balanced (have equal number of instances).

I(Y;C)= Mutual Information

- Mutual information is given as:
  - $I(Y;C) = H(Y) - H(Y|C)$
  - We already know H(Y)
  - H(Y|C) is the entropy of class labels within each cluster, how do we calculate this??

Mutual Information tells us the reduction in the entropy of class labels that we get if we know the cluster labels. (Similar to Information gain in deicison trees)

H(Y|C): conditional entropy of class labels for clustering C

- Consider Cluster-1:
  - P(Y=1|C=1)=3/10 (three triangles in cluster-1)
  - P(Y=2|C=1)=3/10 (three rectangles in cluster-1)
  - P(Y=3|C=1)=4/10 (four stars in cluster-1)
  - Calculate conditional entropy as:

$$H(Y|C=1) = -P(C=1) \sum_{y \in \{1,2,3\}} P(Y=y|C=1) \log(P(Y=y|C=1))$$

$$= -\frac{1}{2} \times [\frac{3}{10} \log\left(\frac{3}{10}\right) + \frac{3}{10} \log(\frac{3}{10}) + \frac{4}{10} \log(\frac{4}{10})] = 0.7855$$

H(Y|C): conditional entropy of class labels for clustering C

- Now, consider Cluster-2:
  - P(Y=1|C=2)=2/10 (two triangles in cluster-1)
  - P(Y=2|C=2)=7/10 (seven rectangles in cluster-1)
  - P(Y=3|C=2)=1/10 (one star in cluster-1)
  - Calculate conditional entropy as:

$$H(Y|C=2) = -P(C=2) \sum_{y \in \{1,2,3\}} P(Y=y|C=2) \log(P(Y=y|C=2))$$

$$= -\frac{1}{2} \times [\frac{2}{10} \log\left(\frac{2}{10}\right) + \frac{7}{10} \log(\frac{7}{10}) + \frac{1}{10} \log(\frac{1}{10})] = \color{red}{0.5784}$$

I(Y;C)

- Finally the mutual information is:

$$\begin{aligned} I(Y;C) &= H(Y) - H(Y|C) \\ &= 1.5 - [0.7855 + 0.5784] \\ &= 0.1361 \end{aligned}$$

The NMI is therefore,

$$NMI(Y,C) = \frac{2 \times I(Y;C)}{[H(Y) + H(C)]}$$

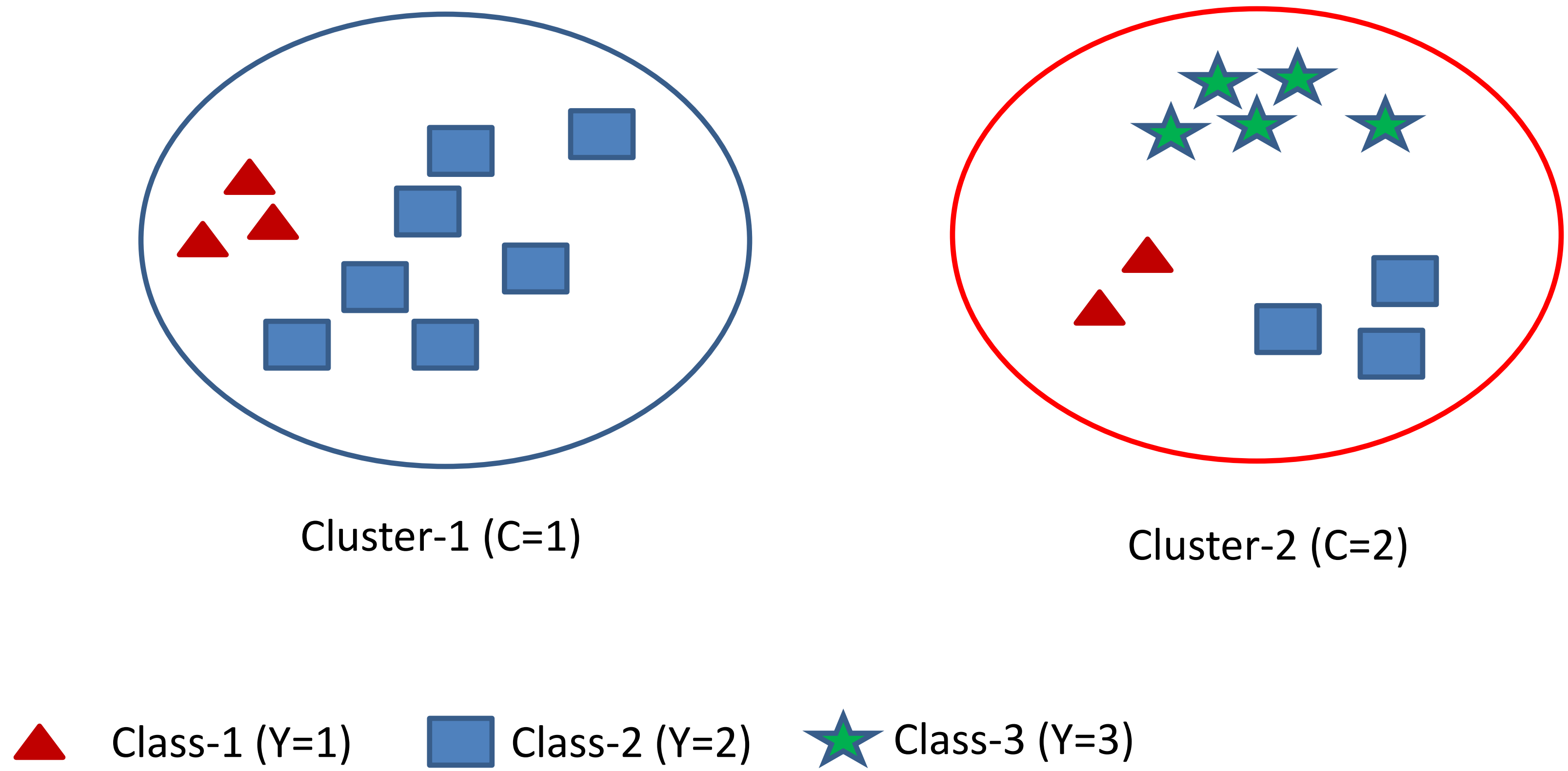
$$NMI(Y,C) = \frac{2 \times 0.1361}{[1.5 + 1]} = 0.1089$$

NMI

- NMI is a good measure for determining the quality of clustering.
- It is an external measure because we need the class labels of the instances to determine the NMI.
- Since it’s normalized we can measure and compare the NMI between different clusterings having different number of clusters.

NMI for Clustering

- Calculate the NMI:



H(Y|C): conditional entropy of class labels for clustering C

- Consider Cluster-1:
  - P(Y=1|C=1)=3/10 (three triangles in cluster-1)
  - P(Y=2|C=1)=7/10 (seven rectangles in cluster-1)
  - P(Y=3|C=1)=0/10 (no stars in cluster-1)
  - Calculate conditional entropy as:

$$H(Y|C=1) = -P(C=1) \sum_{y \in \{1,2,3\}} P(Y=y|C=1) \log(P(Y=y|C=1))$$

$$= -\frac{1}{2} \times [\frac{3}{10} \log\left(\frac{3}{10}\right) + \frac{0}{10} \log(\frac{0}{10}) + \frac{7}{10} \log(\frac{7}{10})] = 0.4406$$

We used 0log(0)=0

H(Y|C): conditional entropy of class labels for clustering C

- Now, consider Cluster-2:
  - P(Y=1|C=2)=2/10 (two triangles in cluster-1)
  - P(Y=2|C=2)=3/10 (three rectangles in cluster-1)
  - P(Y=3|C=2)=5/10 (five stars in cluster-1)
  - Calculate conditional entropy as:

$$H(Y|C=2) = -P(C=2)\sum_{y\in\{1,2,3\}} P(Y=y|C=2)\log(P(Y=y|C=2)\,)$$

$$= -\frac{1}{2}\times[\frac{2}{10}\log\left(\frac{2}{10}\right)+\frac{3}{10}\log(\frac{3}{10})+\frac{5}{10}\log(\frac{5}{10})] = 0.7427$$

I(Y;C)

- Finally the mutual information is:

$$\begin{aligned} I(Y;C) &= H(Y) - H(Y|C) \\ &= 1.5 - [0.4406 + 0.7427] \\ &= 0.3167 \end{aligned}$$

The NMI is therefore,

$$NMI(Y,C) = \frac{2 \times I(Y;C)}{[H(Y) + H(C)]}$$

$$NMI(Y,C) = \frac{2 \times 0.3167}{[1.5 + 1]} = 0.2533$$

Comments

- NMI for the second clustering is higher than the first clustering. It means we would prefer the second clustering over the first.
  - You can see that one of the clusters in the second case contains all instances of class-3 (stars).
- If we have to compare two clustering that have different number of clusters we can still use NMI.