[Normalized Mutual Information] ①

→ It is one of the clustering quality parameter:

Normalized Mutual Information:

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

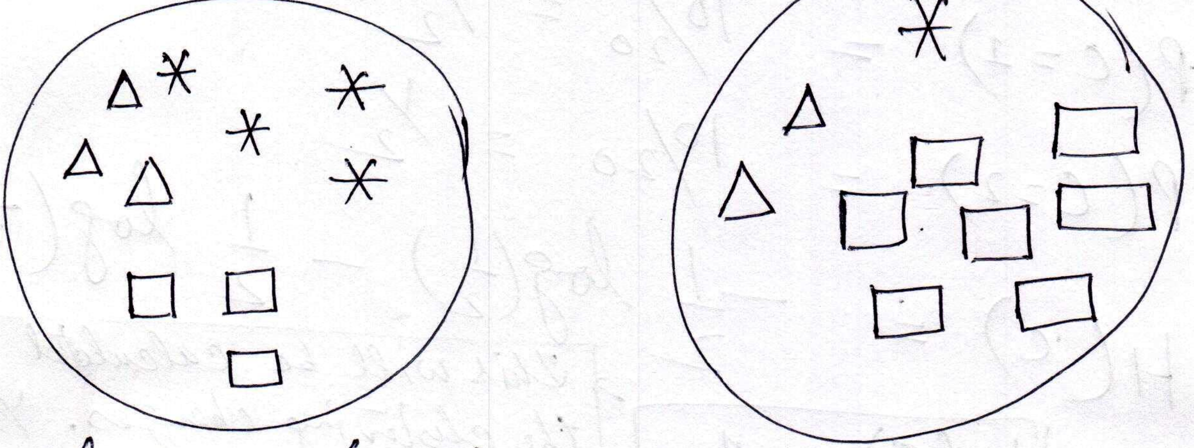
Where,

1) Y = class labels
2) C = cluster labels
3) $H(\cdot)$ = Entropy
4) $I(Y;C)$ = Mutual Information between Y and C.

Note: All logs are base-2.

Calculating NMI for clustering:

⇒ Assume $m = 3$ classes and $k = 2$ clusters ($\triangle$, $\square$, $*$)



cluster-1 ($c=1$) cluster-2 ($c=2$)

$H(Y)$ = Entropy of class labels

$\Rightarrow P(Y=1)\triangle = 5/20 = 1/4$ (class-1: triangle)

$\Rightarrow P(Y=2)\bigstar = 5/20 = 1/4$ (class 2: (star))

$\Rightarrow P(Y=3)\square = 10/20 = 1/2$ (class 3: rectangle)

$$H(Y) = -\sum_{i=1}^{3} P_i \log_2 P_i = -P_1 \log_2 P_1 - P_2 \log_2 P_2 - P_3 \log_2 P_3$$

$$= -\frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)$$

$$\boxed{H(Y) = 1.5}$$

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

$H(C)$ = Entropy of cluster labels

$P(C=1) = 10/20 = 1/2$

$P(C=2) = 10/20 = 1/2$

$$H(C) = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right)$$

$$\boxed{H(C) = 1}$$

This will be calculated every time the clustering changes. You can see from the example that the clusters are balanced i.e have equal no of objects

$I(Y;C)$ = Mutual Information (2)

⇒ Mutual Information is given as:

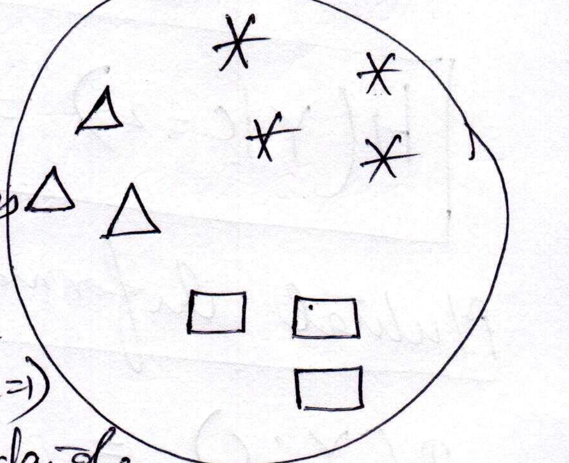
$$I(Y:C) = H(Y) - H(Y|C)$$

⇒ we already know $(H(Y))$

⇒ $H(Y|C)$ is the entropy of class labels within each cluster

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 the decision tree).

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



cluster-1 (C=1)

<u>cluster-1</u>

⇒ $P(Y=1|C=1) = 3/10$ (three tringles in cluster-1)

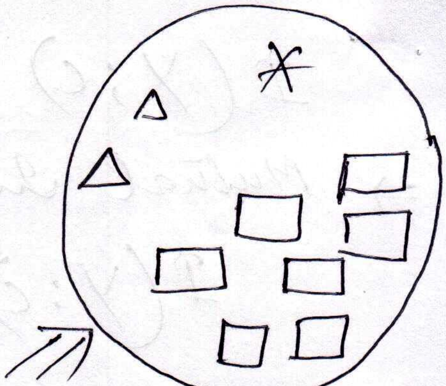
⇒ $P(Y=2|C=1) = 4/10$ (~~Three~~ four stars in $C=1$)

⇒ $P(Y=3|C=1) = 3/10$ (Three rectangles in $c=1$)

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

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

$$\boxed{H(Y|c=1) = 0.7855}$$



Now for cluster-2:

$\Rightarrow P(Y=1|c=2) = \frac{2}{10}$ (Two triangles in $c=2$)

$\Rightarrow P(Y=2|c=2) = \frac{1}{10}$ (One star in $c=2$)

$\Rightarrow P(Y=3|c=2) = 7/10$ (Seven Rectangles in $c=2$)

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

$$= -\frac{1}{2} \times \left[\frac{2}{10} \log \frac{2}{10} + \frac{1}{10} \log \frac{1}{10} + \frac{7}{10} \log \frac{7}{10}\right]$$

$$\boxed{H(Y|c=2) = 0.5784}$$

Mutual Information $I(Y;C)$ is:

$$I(Y;C) = H(Y) - H(Y|C)$$

$$= 1.5 - [0.7855 + 0.5784]$$

$$\boxed{I(Y;C) = 0.1361}$$

Therefore NMI is: ③

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

$$= \frac{2 \times 0.1361}{[1.5 + 1]}$$

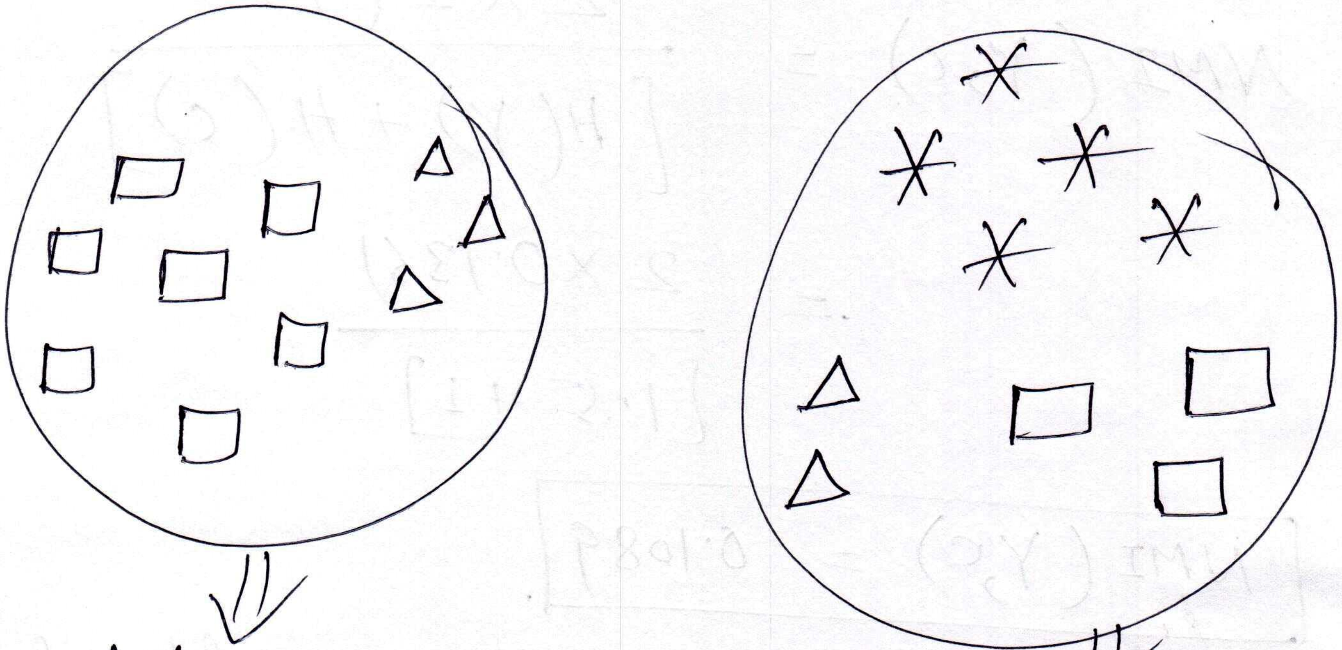
$$\boxed{NMI(Y, C) = 0.1089}$$

⇒ NMI is a good measure of quality of clustering.

⇒ It is an external measure because we need the class labels of the instances to determine the NMI.

⇒ Since it is normalized we can measure and compare the NMI between different clusterings having different number of clusters.

Excercise



Cluster-1

Cluster-2

Class labels : $\left(\overset{Y=1}{\triangle}, \overset{Y=2}{\ast}, \overset{Y=3}{\square} \right)$

$$\boxed{NMI(Y, C) = 0.2533}$$

So If we see two clusting approaches here the second clustering is better because NMI is higher.

0.1089 < 0.2533

So we will use the second approach of clustering.