

# Information mutuelle



# Dans les épisodes précédents...

- Notion de valeur / prix de l'information :
  - L'information est-elle facile à obtenir ?
  - Y a-t-il beaucoup de possibilités et d'incertitude ?
  - Mesure de **quantité d'information** (relative à un évènement) :  $I = \log_2(N/n)$



# Dans les épisodes précédents...

- Notion de valeur / prix de l'information :

- L'information est-elle facile à obtenir ?
- Y a-t-il beaucoup de possibilités et d'incertitude ?
- Mesure de **quantité d'information** (relative à un évènement) :  $I = \log_2(N/n)$



- Quantité moyenne d'information dans une distribution :

- **Entropie**
- $H(I) = -\sum_{i \in I} p_i \log_2(p_i)$



# Dans les épisodes précédents...

- Notion de valeur / prix de l'information :

- L'information est-elle facile à obtenir ?
- Y a-t-il beaucoup de possibilités et d'incertitude ?
- Mesure de **quantité d'information** (relative à un évènement) :  $I = \log_2(N/n)$



- Quantité moyenne d'information dans une distribution :

- **Entropie**

- $H(I) = -\sum_{i \in I} p_i \log_2(p_i)$



- Dépendance entre variables, et **information** partagée / **mutuelle** :

- $IM(C, T) = H(C) + H(T) - H(C, T)$

- $IM(C, T) = H(T) - H(T|C) = H(C) - H(C|T)$



# Calcul de l'information mutuelle

Comment calcule-t-on l'information mutuelle en pratique ?

$$IM(X, Y) = H(Y) - H(Y|X)$$

$$IM(X, Y) = H(Y) - H(X, Y) + H(X)$$

# Calcul de l'information mutuelle

$$IM(X, Y) = H(Y) - H(X, Y) + H(X)$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x p_x \log_2 p_x$$

# Calcul de l'information mutuelle

$$IM(X, Y) = H(Y) - H(X, Y) + H(X)$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x \textcolor{red}{p_x} \log_2 p_x$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x \sum_{\textcolor{red}{y}} \textcolor{red}{p_{x,y}} \log_2 p_x$$

# Calcul de l'information mutuelle

$$IM(X, Y) = H(Y) - H(X, Y) + H(X)$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x p_x \log_2 p_x$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x \sum_y p_{x,y} \log_2 p_x$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$



# Zoom sur une propriété de l'information mutuelle

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$

# Zoom sur une propriété de l'information mutuelle

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$

donc

$$IM(X, Y) = - \sum_{x,y} p_{x,y} \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$

# Zoom sur une propriété de l'information mutuelle

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$

donc

$$IM(X, Y) = - \sum_{x,y} p_{x,y} \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$

$$IM(X, Y) = \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x p_y}$$

# Zoom sur une propriété de l'information mutuelle

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$

donc

$$IM(X, Y) = - \sum_{x,y} p_{x,y} \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$

$$IM(X, Y) = \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x p_y} = D(p_{x,y} \parallel p_x p_y)$$



Divergence de Kullback-Leibler

# Divergence de Kullback-Leibler

- Définition :

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

Mesure l'écart / l'**information de discrimination** entre les distributions de probabilités  $q$  et  $p$ .

# Divergence de Kullback-Leibler

- Définition :

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

Mesure l'écart / l'**information de discrimination** entre les distributions de probabilités  $q$  et  $p$ .

- Lien avec l'information mutuelle :

$$IM(X, Y) = D(p_{x,y} \parallel p_x p_y)$$

- ✓ Mesure l'écart entre la distribution jointe et le produit des marginales.
- ✓  $IM(X, Y) = H(Y) + H(X) - H(X, Y)$

Reprenons le calcul de l'information mutuelle...

$$IM(X, Y) = H(Y) - H(X, Y) + H(X)$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x p_x \log_2 p_x$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x \sum_y p_{x,y} \log_2 p_x$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$

Reprenons le calcul de l'information mutuelle...

$$IM(X, Y) = H(Y) - H(X, Y) + H(X)$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x p_x \log_2 p_x$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x \sum_y p_{x,y} \log_2 p_x$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_x p_x \sum_y p_{y|x} \log_2 p_{y|x}$$

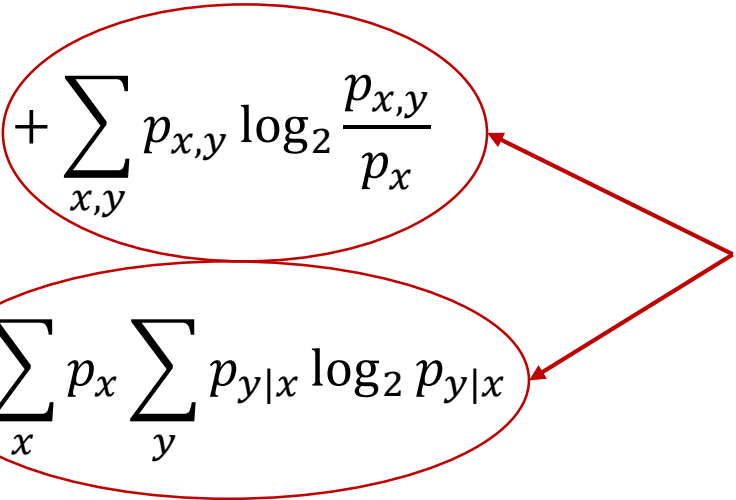


Reprenons le calcul de l'information mutuelle...

$$IM(X, Y) = H(Y) - H(X, Y) + H(X)$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x p_x \log_2 p_x$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 p_{x,y} - \sum_x \sum_y p_{x,y} \log_2 p_x$$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x}$$


$-H(Y|X)$

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_x p_x \sum_y p_{y|x} \log_2 p_{y|x}$$

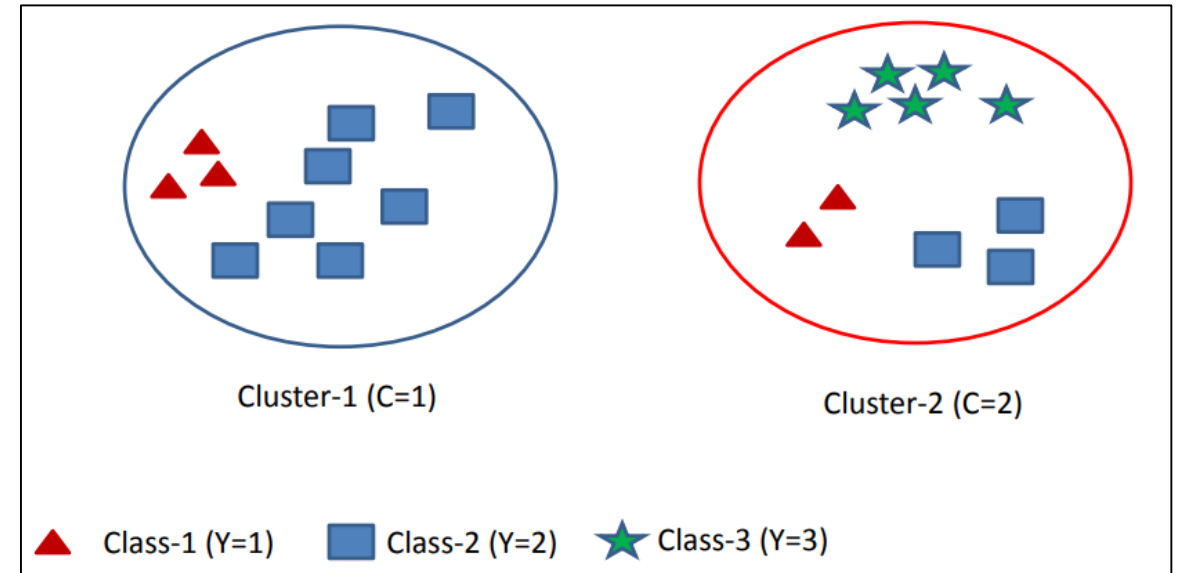
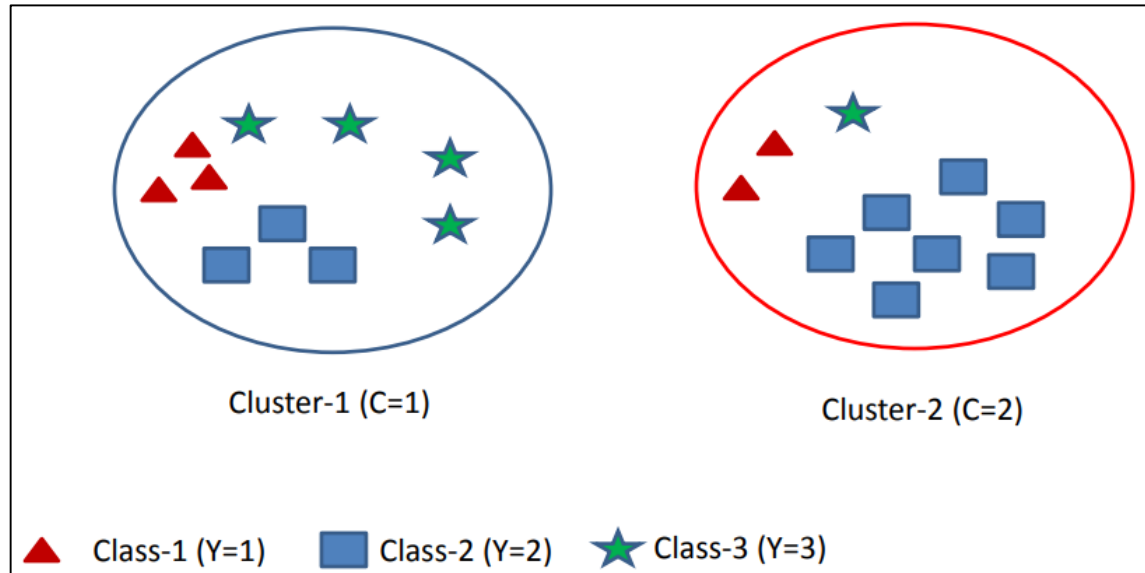
# Calcul de l'information mutuelle

Mise en pratique avec l'équation :

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_x p_x \sum_y p_{y|x} \log_2 p_{y|x}$$

# Calcul de l'information mutuelle

Exemple d'application : clustering



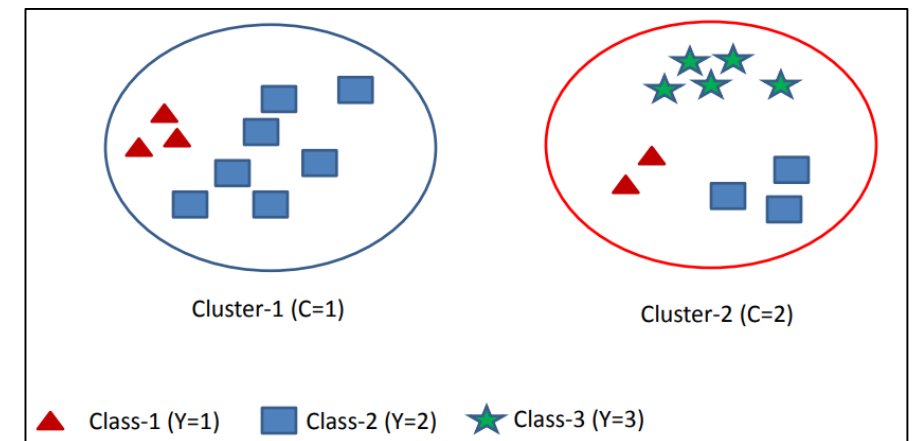
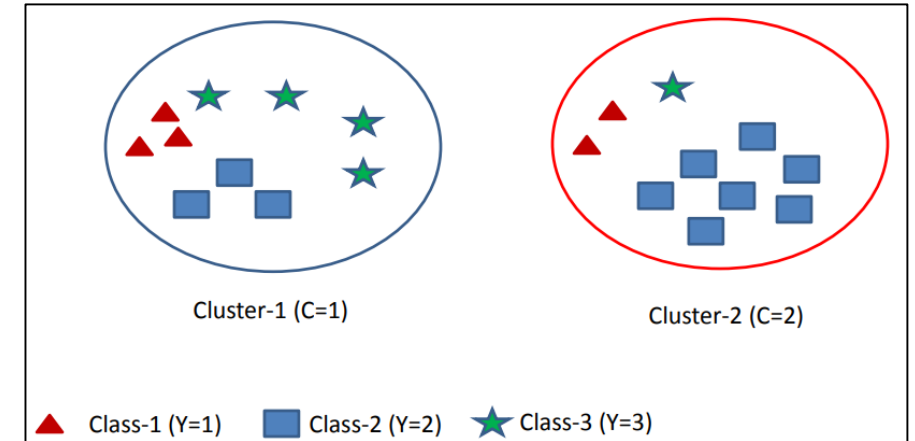
Comment choisir le meilleur clustering ?

# Calcul de l'information mutuelle

Meilleur clustering :

Maximise l'information mutuelle entre Y et C

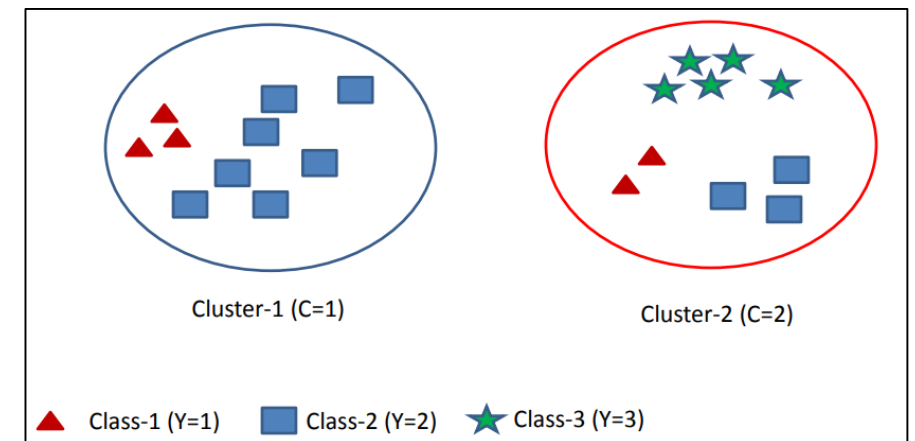
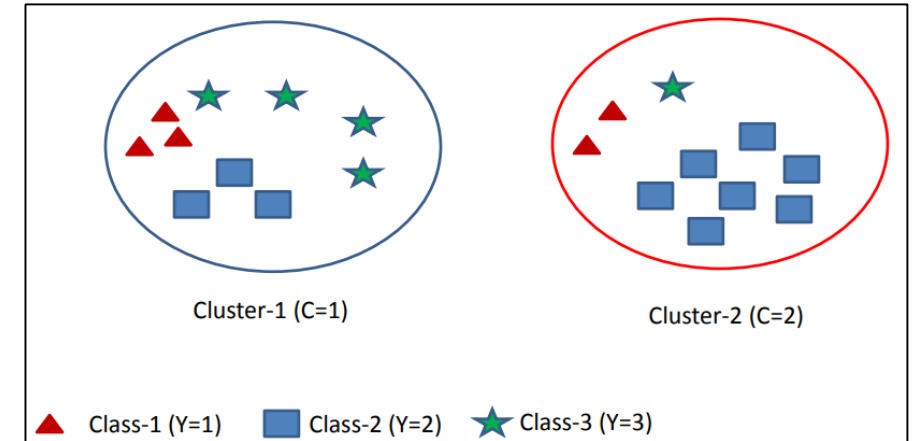
$$IM(Y, C)_{\text{bon clustering}} > IM(Y, C)_{\text{mauvais clustering}}$$



# Calcul de l'information mutuelle

$$IM(C, Y) = - \sum_y p_y \log_2 p_y + \sum_c p_c \sum_y p_{y|c} \log_2 p_{y|c}$$

Information mutuelle pour le premier clustering ?



# Calcul de l'information mutuelle

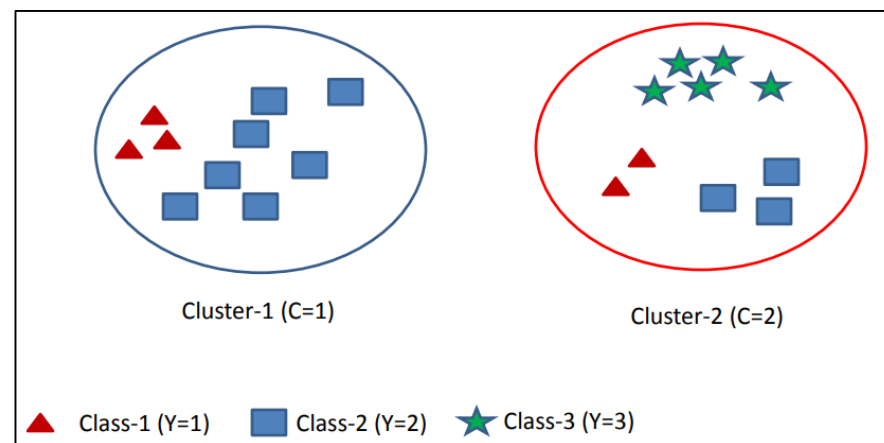
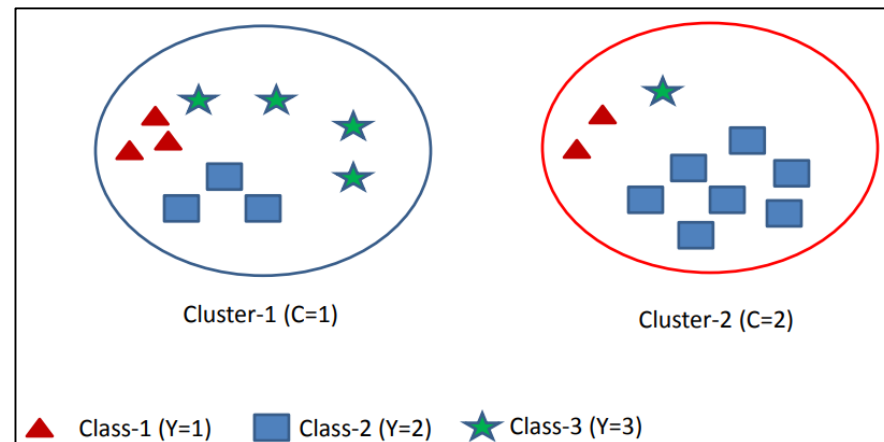
$$IM(Y, C) = - \sum_y p_y \log_2 p_y + \sum_c p_c \sum_y p_{y|c} \log_2 p_{y|c}$$

$$P(y = 1) = \frac{5}{20} = \frac{1}{4}$$

$$P(y = 2) = \frac{10}{20} = \frac{1}{2}$$

$$P(y = 3) = \frac{5}{20} = \frac{1}{4}$$

$$H(Y) = -\frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{4} \log_2 \frac{1}{4} = \frac{3}{2} = 1.5$$



# Calcul de l'information mutuelle

$$IM(Y, C) = - \sum_y p_y \log_2 p_y + \sum_c p_c \sum_y p_{y|c} \log_2 p_{y|c}$$

Pour  $c = 1$  :

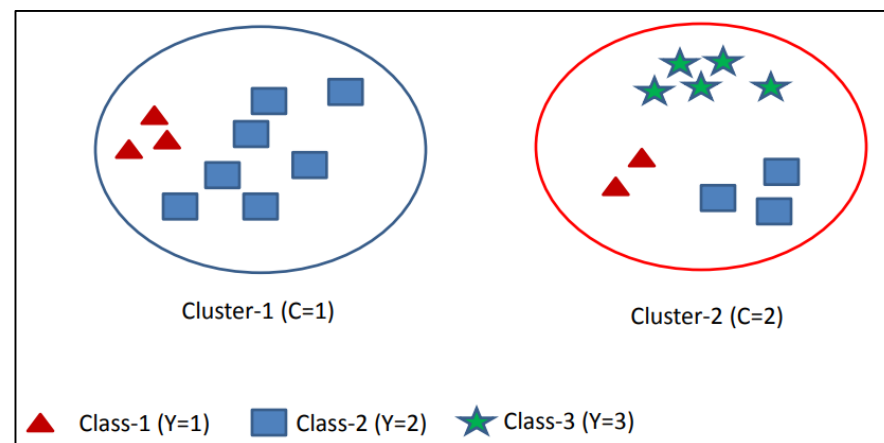
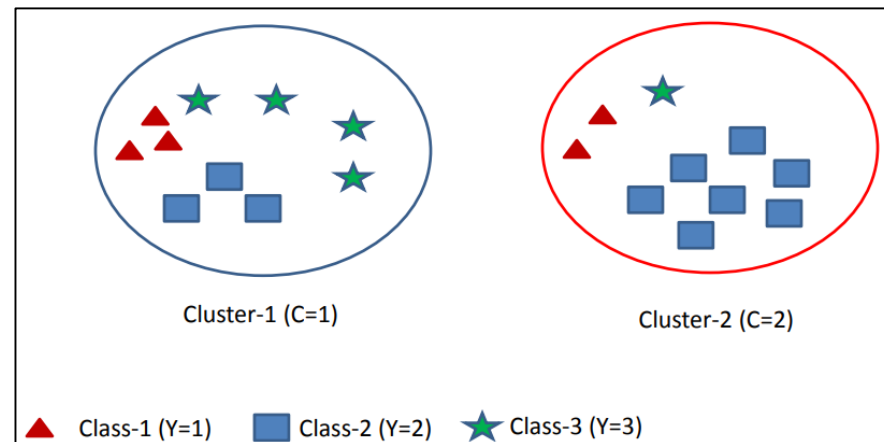
$$P(y = 1|c = 1) = \frac{3}{10}$$

$$P(y = 2|c = 1) = \frac{3}{10}$$

$$P(y = 3|c = 1) = \frac{4}{10}$$

$$P(c = 1) = \frac{10}{20} = \frac{1}{2}$$

$$-H(Y|c = 1) = \frac{1}{2} \left[ \frac{3}{10} \log_2 \frac{3}{10} + \frac{3}{10} \log_2 \frac{3}{10} + \frac{4}{10} \log_2 \frac{4}{10} \right] \approx -0.7855$$



# Calcul de l'information mutuelle

$$IM(Y, C) = - \sum_y p_y \log_2 p_y + \sum_c p_c \sum_y p_{y|c} \log_2 p_{y|c}$$

Pour  $c = 2$  :

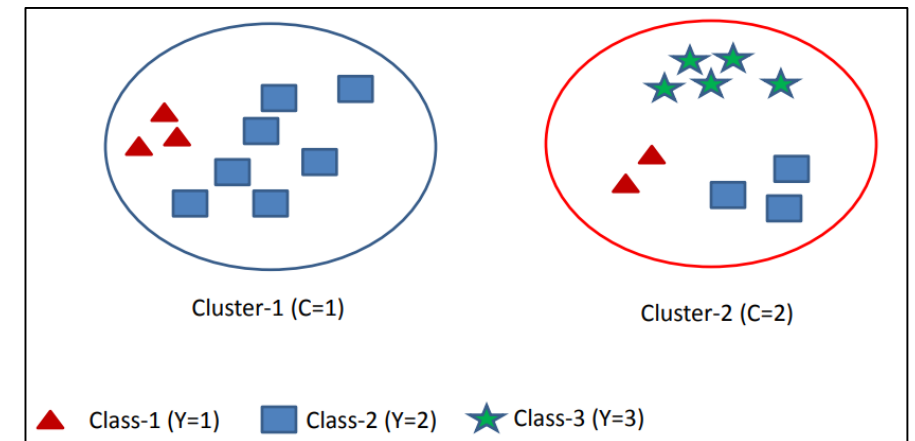
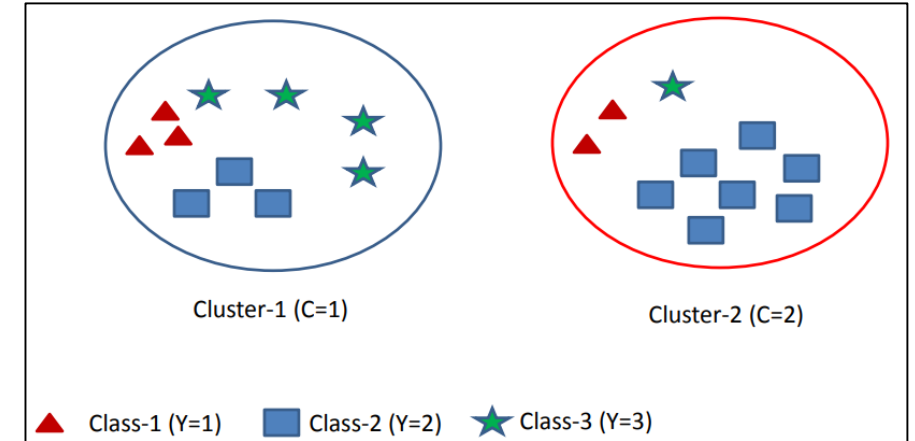
$$P(y = 1|c = 2) = \frac{2}{10}$$

$$P(y = 2|c = 2) = \frac{7}{10}$$

$$P(y = 3|c = 2) = \frac{1}{10}$$

$$P(c = 2) = \frac{10}{20} = \frac{1}{2}$$

$$-H(Y|c = 2) = \frac{1}{2} \left[ \frac{2}{10} \log_2 \frac{2}{10} + \frac{7}{10} \log_2 \frac{7}{10} + \frac{1}{10} \log_2 \frac{1}{10} \right] \approx -0.5784$$



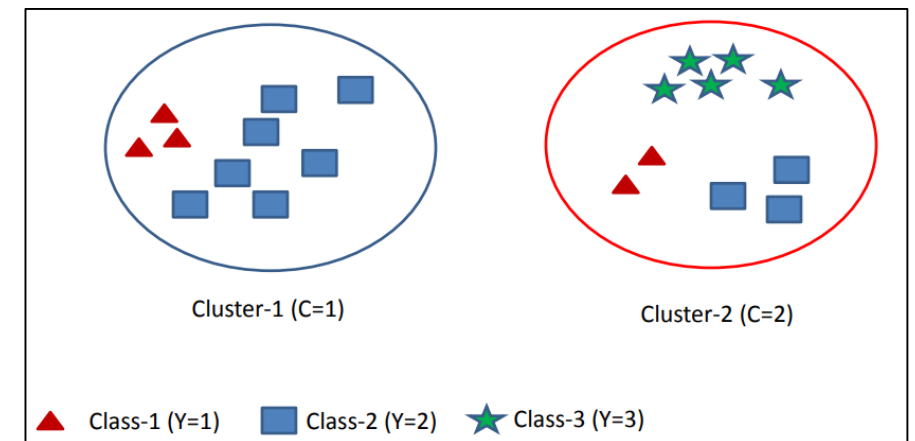
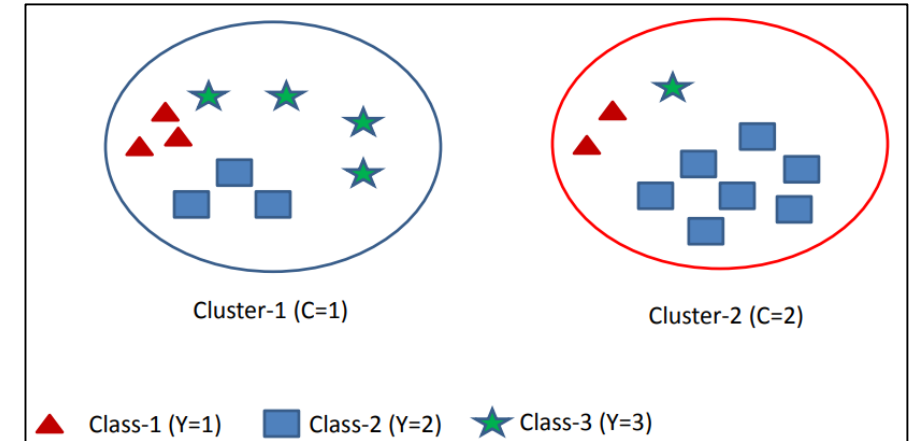


# Calcul de l'information mutuelle

$$IM(C, Y) = - \sum_y p_y \log_2 p_y + \sum_c p_c \sum_y p_{y|c} \log_2 p_{y|c}$$

Information mutuelle pour le premier clustering :

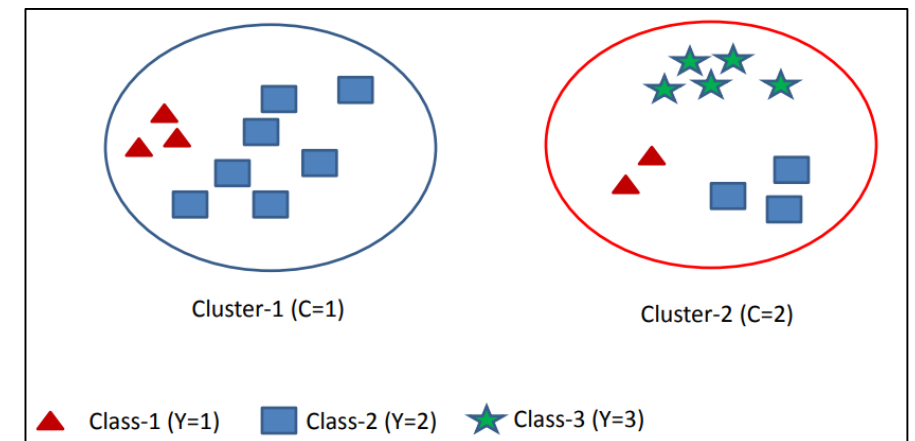
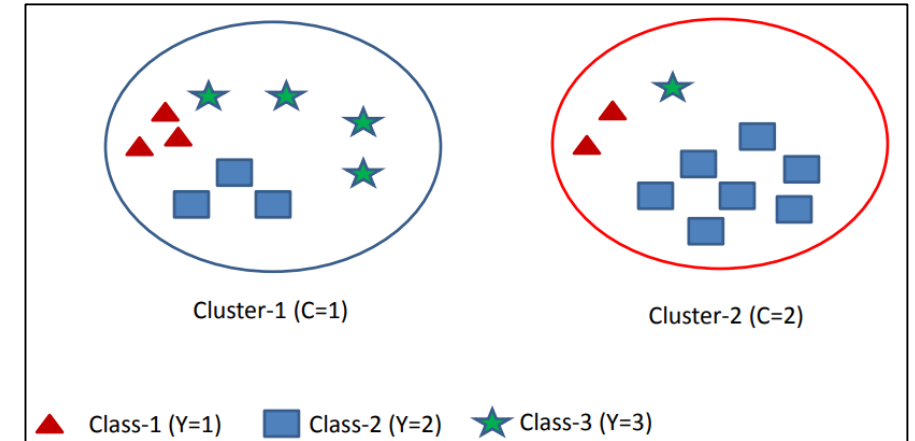
$$IM(C, Y)_{clust\alpha} \approx 1.5 - 0.7855 - 0.5784 = 0.1361$$



# Calcul de l'information mutuelle

$$IM(C, Y) = - \sum_y p_y \log_2 p_y + \sum_c p_c \sum_y p_{y|c} \log_2 p_{y|c}$$

Information mutuelle pour le deuxième clustering ?

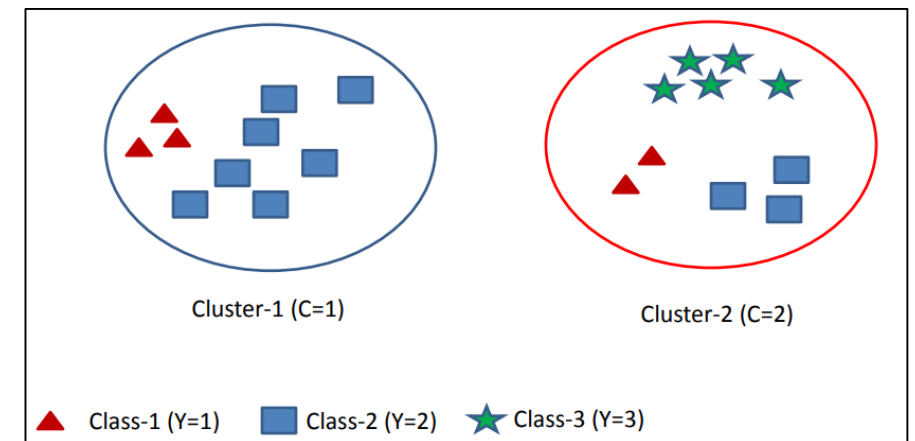
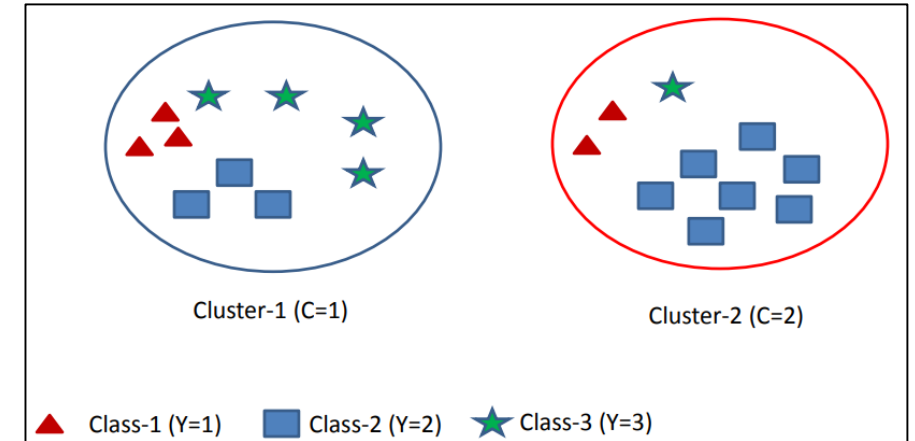


# Calcul de l'information mutuelle

$$IM(C, Y) = - \sum_y p_y \log_2 p_y + \sum_c p_c \sum_y p_{y|c} \log_2 p_{y|c}$$

Information mutuelle pour le deuxième clustering :

$$IM(C, Y)_{clust\beta} \approx 1.5 - 0.4406 - 0.7427 = 0.3167$$



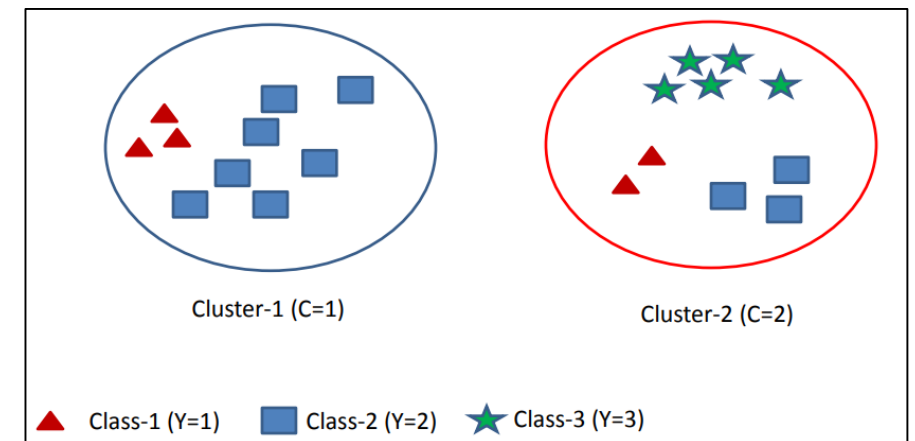
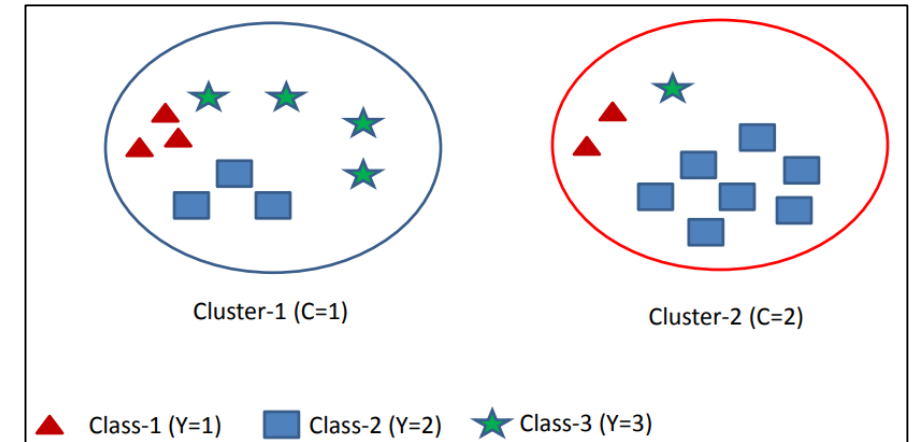
# Calcul de l'information mutuelle

$$IM(C, Y)_{clust\alpha} \approx 0.1361$$

$$IM(C, Y)_{clust\beta} \approx 0.3167$$

Quel est le meilleur clustering ?

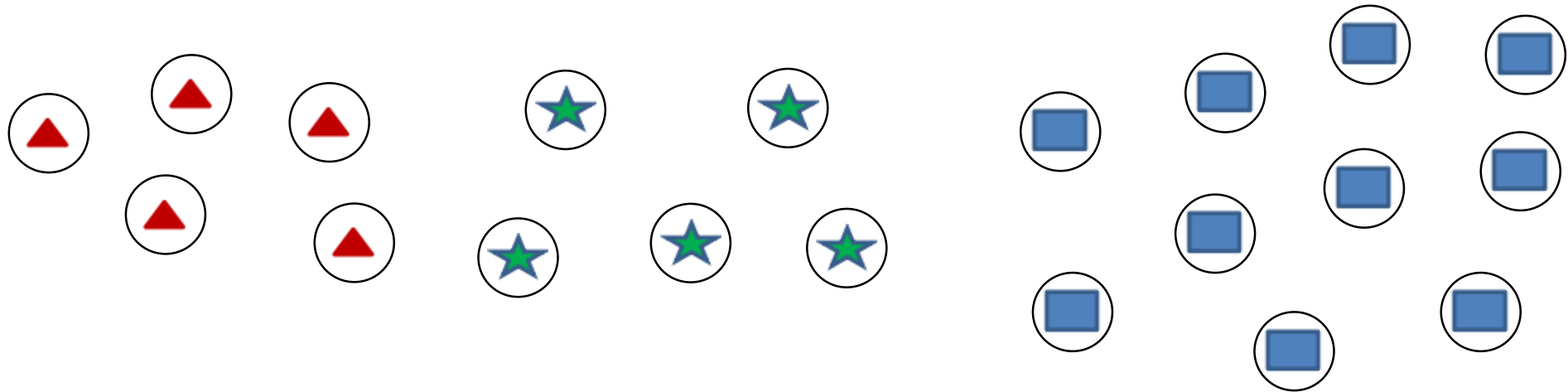
Pourquoi ?



# Calcul de l'information mutuelle

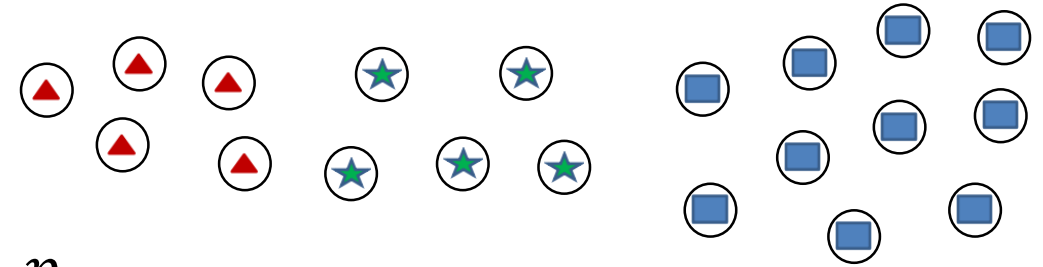
Et si on proposait un 3<sup>ème</sup> clustering avec 20 clusters ?

Information mutuelle pour ce troisième clustering ?



# Calcul de l'information mutuelle

Information mutuelle avec 20 clusters :



$$IM(C, Y) = -\sum_y p_y \log_2 p_y + \sum_c p_c \sum_y p_{y|c} \log_2 p_{y|c}$$

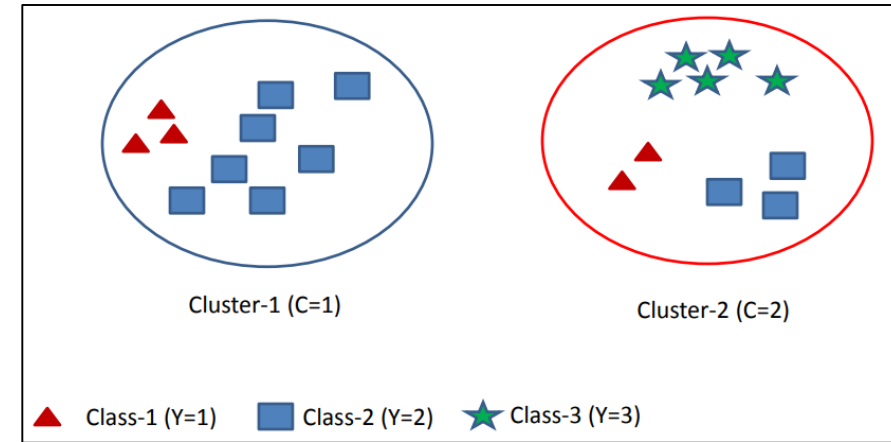
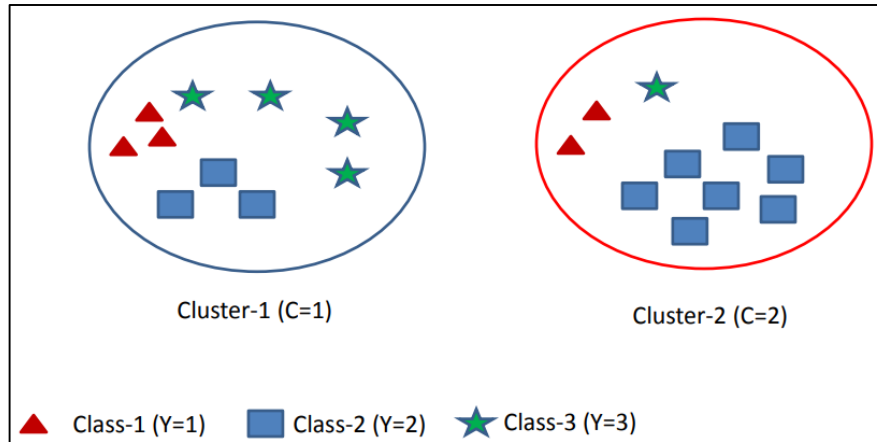
mais  $p_{y|c} = 0$  ou  $1$ , donc :

$$IM(C, Y)_{clust\gamma} = -\sum_y p_y \log_2 p_y = 1.5$$

(même chose pour 3 clusters)

# Calcul de l'information mutuelle

Quel problème avons-nous découvert ?



$$\begin{aligned}IM(C, Y)_{clust\alpha} &\approx 0.1361 \\IM(C, Y)_{clust\beta} &\approx 0.3167 \\IM(C, Y)_{clust\gamma} &= 1.5\end{aligned}$$

# Information mutuelle normalisée

Normalisation possible :

$$NMI(C, Y) = \frac{2 * MI(C, Y)}{H(C) + H(Y)}$$

Cette mesure permet de comparer des clusterings ayant des nombres de clusters différents.



# Information mutuelle normalisée

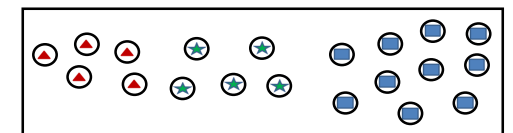
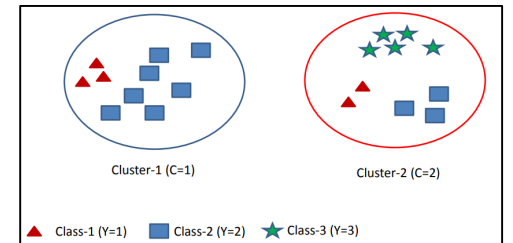
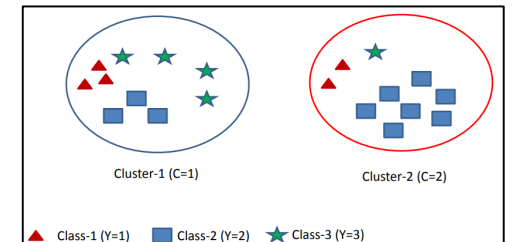
$$NMI(C, Y) = \frac{2 * MI(C, Y)}{H(C) + H(Y)}$$

Nouvelles valeurs ?

$$IM(C, Y)_{clust\alpha} \approx 0.1361$$

$$IM(C, Y)_{clust\beta} \approx 0.3167$$

$$IM(C, Y)_{clust\gamma} = 1.5$$



# Information mutuelle normalisée

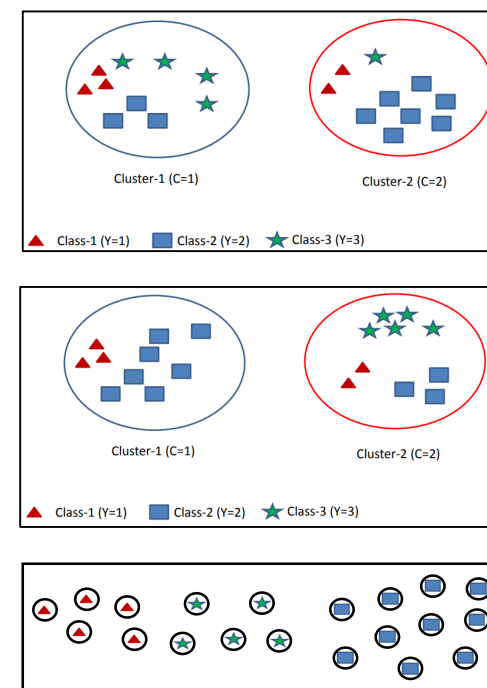
$$NMI(C, Y) = \frac{2 * MI(C, Y)}{H(C) + H(Y)}$$

Pour les clustering  $\alpha$  et  $\beta$  :

$$H(C) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

Pour le clustering  $\gamma$  :

$$H(C) = -20 \frac{1}{20} \log_2 \frac{1}{20} \approx 4.3219$$



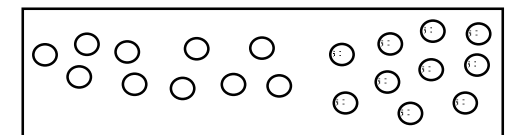
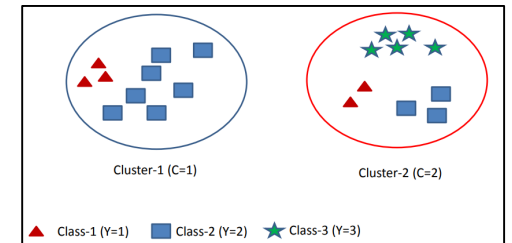
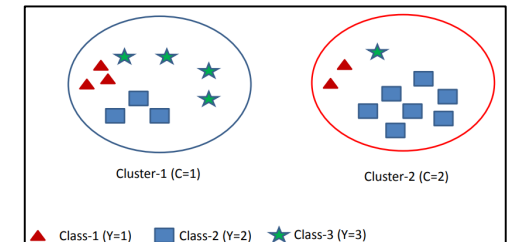
Valeur bien plus grande, un peu disproportionnée, empêche la comparaison !

# Information mutuelle normalisée

$$NMI(C, Y) = \frac{2 * MI(C, Y)}{H(C) + H(Y)}$$

Nouvelles valeurs :

- $NMI(C, Y)_{clust\alpha} \approx \frac{2 * 0.1361}{1 + 1.5} \approx 0.1089$
- $NMI(C, Y)_{clust\beta} \approx 0.2534$
- $NMI(C, Y)_{clust\gamma} \approx 0.5155$



# Information mutuelle normalisée

Il existe en fait plusieurs normalisations possibles, dont :

- Mesure de redondance :  $R = \frac{IM(X,Y)}{H(X)+H(Y)} = 1 - \frac{H(X,Y)}{H(X)+H(Y)}$ , avec  $R = 0$  si  $X$  et  $Y$  indépendants, et  $R_{max} = \frac{\min\{H(X),H(Y)\}}{H(X)+H(Y)}$
- Incertitude symétrique :  $U(X,Y) = 2 * \frac{IM(X,Y)}{H(X)+H(Y)}$  : moyenne harmonique des coefficients d'incertitude  $C_{XY} = \frac{IM(X,Y)}{H(Y)}$  et  $C_{YX} = \frac{IM(X,Y)}{H(X)}$
- Ratio de quantités d'information :  $IQR(X,Y) = \frac{IM(X,Y)}{H(X,Y)} = \frac{H(X)+H(Y)}{H(X,Y)} - 1$  : populaire en traitement d'image
- Normalisation entre 0 et 1 :  $NMI(X,Y) = \frac{IM(X,Y)}{denom}$  où *denom* peut être le min, le max, ou la moyenne arithmétique ou géométrique des entropies. Par exemple :  $NMI(X,Y) = \frac{IM(X,Y)}{\min\{H(X),H(Y)\}}$

# Autre utilisation de l'information mutuelle

Sélection de variables / attributs / features :

Nous voulons prédire le risque qu'un patient ait des complications après une crise cardiaque.

Nous avons récolté plein de données, mais nous ne savons pas lesquelles sont pertinentes :

Age, poids, taille, code postal, couleur des cheveux et des yeux, glycémie, saturation du sang en oxygène, rythme cardiaque au repos et pendant l'effort, etc.

Comment faire le tri ?

# Sélection de variables

Propriété désirée :

Les variables sélectionnées sont liées au risque de complications

Comment faire?

# Sélection des variables associées au risque de complications

Méthode : Maximiser l'IM entre chaque variable et le risque de complications

Age	Poids	Taille	Code postal	Couleur cheveux	Couleur yeux	Glycémie	Saturation	Rythme repos	Rythme effort	Complications
78	84	168	83113	Brun	Bleu	1.02	97	73	113	Oui
67	93	170	83271	Blond	Bleu	1.10	92	82	115	Oui
84	110	186	83230	Chatain	Marron	1.30	90	90	144	Oui
70	68	158	83000	Chatain	Marron	0.75	96	78	136	Non
93	73	182	83018	Roux	Bleu	1.12	89	63	147	Non
59	83	171	83620	Brun	Marron	0.86	100	84	128	Non
80	99	143	83330	Chatain	Bleu	0.96	99	77	139	Non

# Sélection des variables associées au risque de complications

Méthode : Maximiser l'IM entre chaque variable et le risque de complication

1. Calculez l'IM entre chaque variable (tranches d'âge/de poids...) et le risque de complications
2. Sélectionnez les variables les plus pertinentes

Age	Poids	Taille	Code postal	Couleur cheveux	Couleur yeux	Glycémie	Saturation	Rythme repos	Rythme effort	Complications
78	84	168	83113	Brun	Bleu	1.02	97	73	113	Oui
67	93	170	83271	Blond	Bleu	1.10	92	82	115	Oui
84	110	186	83230	Chatain	Marron	1.30	90	90	144	Oui
70	68	158	83000	Chatain	Marron	0.75	96	78	136	Non
93	73	182	83018	Roux	Bleu	1.12	89	63	147	Non
59	83	171	83620	Brun	Marron	0.86	100	84	128	Non
80	99	143	83330	Chatain	Bleu	0.96	99	77	139	Non



# Sélection des variables associées au risque de complications

Définition de variables aléatoires discrètes :

$age \in \{[0,70] ; ]70, \infty[ \}$

$poids \in \{[0,75] ; ]75,85] ; ]85, \infty[ \}$

$taille \in \{[0,160] ; ]160,180]; ]180, \infty[ \}$

$code \in \{830..; 831..; 832..; 833..; 836. \}$

$cheveux \in \{Brun ; Bond ; Chatain ; Roux \}$

$yeux \in \{Bleu, Marron \}$

$glycémie \in \{[0,1.10[ ; [1.10, \infty[ \}$

$saturation \in \{[0,95[ ; [95, \infty[ \}$

$repos \in \{[0,85] ; ]85, \infty[ \}$

$effort \in \{[0,140] ; ]140, \infty[ \}$

Age	Poids	Taille	Code postal	Couleur cheveux	Couleur yeux	Glycémie	Sat.	Rythme repos	Rythme effort	Complications
78	84	168	83113	Brun	Bleu	1.02	97	73	113	Oui
67	93	170	83271	Blond	Bleu	1.10	92	82	115	Oui
84	110	186	83230	Chatain	Marron	1.30	90	90	144	Oui
70	68	158	83000	Chatain	Marron	0.75	96	78	136	Non
93	73	182	83018	Roux	Bleu	1.12	89	63	147	Non
59	83	171	83620	Brun	Marron	0.86	100	84	128	Non
80	99	143	83330	Chatain	Bleu	0.96	99	77	139	Non

Pourquoi discrétise-t-on les valeurs que peuvent prendre ces variables ?

$$IM(X, Y) = - \sum_y p_y \log_2 p_y + \sum_x p_x \sum_y p_{y|x} \log_2 p_{y|x}$$

$$NMI(X, Y) = \frac{2 * MI(X, Y)}{H(X) + H(Y)}$$

# Sélection des variables associées au risque de complications

Définition de variables aléatoires discrètes :

$age \in \{[0,70] ; ]70, \infty[ \}$

$poids \in \{[0,75] ; ]75,85] ; ]85, \infty[ \}$

$taille \in \{[0,160] ; ]160,180]; ]180, \infty[ \}$

$code \in \{830..; 831..; 832..; 833..; 836. \}$

$cheveux \in \{Brun ; Bond ; Chatain ; Roux \}$

$yeux \in \{Bleu, Marron \}$

$glycémie \in \{[0,1.10[ ; [1.10, \infty[ \}$

$saturation \in \{[0,95[ ; [95, \infty[ \}$

$repos \in \{[0,85] ; ]85, \infty[ \}$

$effort \in \{[0,140] ; ]140, \infty[ \}$

Age	Poids	Taille	Code postal	Couleur cheveux	Couleur yeux	Glycémie	Sat.	Rythme repos	Rythme effort	Complications
78	84	168	83113	Brun	Bleu	1.02	97	73	113	Oui
67	93	170	83271	Blond	Bleu	1.10	92	82	115	Oui
84	110	186	83230	Chatain	Marron	1.30	90	90	144	Oui
70	68	158	83000	Chatain	Marron	0.75	96	78	136	Non
93	73	182	83018	Roux	Bleu	1.12	89	63	147	Non
59	83	171	83620	Brun	Marron	0.86	100	84	128	Non
80	99	143	83330	Chatain	Bleu	0.96	99	77	139	Non

Exercice : calculer  $IM(Complications; Age)$  et  $NMI(Complications; Age)$

# Sélection des variables associées au risque de complications

Age :

$$IM(Complications; Age) \approx 0.0202$$

$$NMI(Complications; Age) \approx 0.0205$$

Poids :

$$IM(Complications; Poids) \approx 0.3060$$

$$NMI(Complications; Poids) \approx 0.2408$$

Glycémie :

$$IM(Complications; Glycémie) \approx 0.1280$$

$$NMI(Complications; Glycémie) \approx 0.1299$$

etc...

Age	Poids	Taille	Code postal	Couleur cheveux	Couleur yeux	Glycémie	Sat.	Rythme repos	Rythme effort	Complications
78	84	168	83113	Brun	Bleu	1.02	97	73	113	Oui
67	93	170	83271	Blond	Bleu	1.10	92	82	115	Oui
84	110	186	83230	Chatain	Marron	1.30	90	90	144	Oui
70	68	158	83000	Chatain	Marron	0.75	96	78	136	Non
93	73	182	83018	Roux	Bleu	1.12	89	63	147	Non
59	83	171	83620	Brun	Marron	0.86	100	84	128	Non
80	99	143	83330	Chatain	Bleu	0.96	99	77	139	Non

Quelle(s) variable(s) est-il raisonnable d'utiliser ?

# Sélection des variables associées au risque de complications

Age :

$$IM(Complications; Age) \approx 0.0202$$

$$NMI(Complications; Age) \approx 0.0205$$

Poids :

$$IM(Complications; Poids) \approx 0.3060$$

$$NMI(Complications; Poids) \approx 0.2408$$

Glycémie :

$$IM(Complications; Glycémie) \approx 0.1280$$

$$NMI(Complications; Glycémie) \approx 0.1299$$

etc...

Méthode dite « greedy » :

1. On sélectionne la meilleure variable
2. On sélectionne la deuxième meilleure variable
3. etc.
4. ... jusqu'à ce qu'on ait assez de variables :
  - gain d'information négligeable
  - ou
  - amélioration de la prédiction négligeable

Remarque : il existe d'autres algorithmes pour optimiser la sélection

Information mutuelle entre plus de deux variables

