Exercice 1

a) Le risque empirique pour l'EST_1 :

pred

6	8
1	2

$$\frac{1}{17}(1\times 8 + x \times 1) = \frac{8+x}{17}$$

Le risque empirique pour l'EST_2 =

pred

4	1
3	9

$$\frac{1}{17}(1\times 1 + x \times 3) = \frac{1+3x}{17}$$

$$\frac{8}{17} + \frac{x}{17} \geqslant \frac{1}{17} + \frac{3x}{17}$$

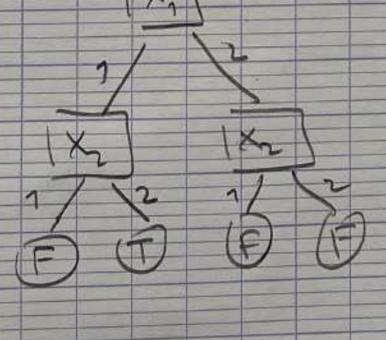
$$\Leftrightarrow \frac{7}{17} \geqslant \frac{2x}{17}$$

$$\Leftrightarrow \boxed{\frac{7}{2} \geqslant x}$$

si $x < 3{,}5 \Rightarrow$ EST_2 est avantagé

$x \geqslant 3{,}5 \Rightarrow$ EST_1 est avantagé

b)



$$\frac{1}{17} \times (3\times 2 + 1\times 1) = \frac{5}{17}$$

c) L'arbre de décision nous permet d'avoir des résultats optimaux

d) Pour y = True

	1	2
x_1	5	2
x_2	2	5

Pour y = False

	1	2
x_1	3	7
x_2	6	4

c) Pour y = True

(x_1, x_2)

$(1,1)$ $\frac{7}{17} \times \frac{5}{7} \times \frac{2}{7} = \frac{10}{119}$

$(1,2)$ $\frac{7}{17} \times \frac{5}{7} \times \frac{5}{7} = \frac{25}{119}$

$(2,1)$ $\frac{7}{17} \times \frac{2}{7} \times \frac{2}{7} = \frac{4}{119}$

$(2,2)$ $\frac{7}{17} \times \frac{2}{7} \times \frac{5}{7} = \frac{10}{119}$

Pour y = False

(x_1, x_2)

$(1,1)$ $\frac{10}{17} \times \frac{3}{10} \times \frac{6}{10} = \frac{9}{85}$

$(1,2)$ $\frac{10}{17} \times \frac{3}{10} \times \frac{4}{10} = \frac{6}{85}$

$(2,1)$ $\frac{10}{17} \times \frac{7}{10} \times \frac{6}{10} = \frac{21}{85}$

$(2,2)$ $\frac{10}{17} \times \frac{7}{10} \times \frac{4}{10} = \frac{14}{85}$

	1	2
1	F	T
2	F	F

Exercice 2

a) Voir le fichier 'question2)a.png'

On a les 9 premiers points en rouge et les 5 derniers en bleu.

On peut utiliser HCA ou KMeans pour distinguer les deux classes.

b) Voir 'dendrogram.png'.

Exercice 3

a) Pour savoir si les deux variables concernées sont indépendantes, on peut utiliser le test χ^2.

On calcule d'abord les effectifs noirés espérés avec # étant le nombre d'évènements :

$$E_{i,j} = \frac{1}{n} \sum \#\{x = i\} \#\{y = j\}$$

Ensuite, il faut calculer les effectifs noirés observés :

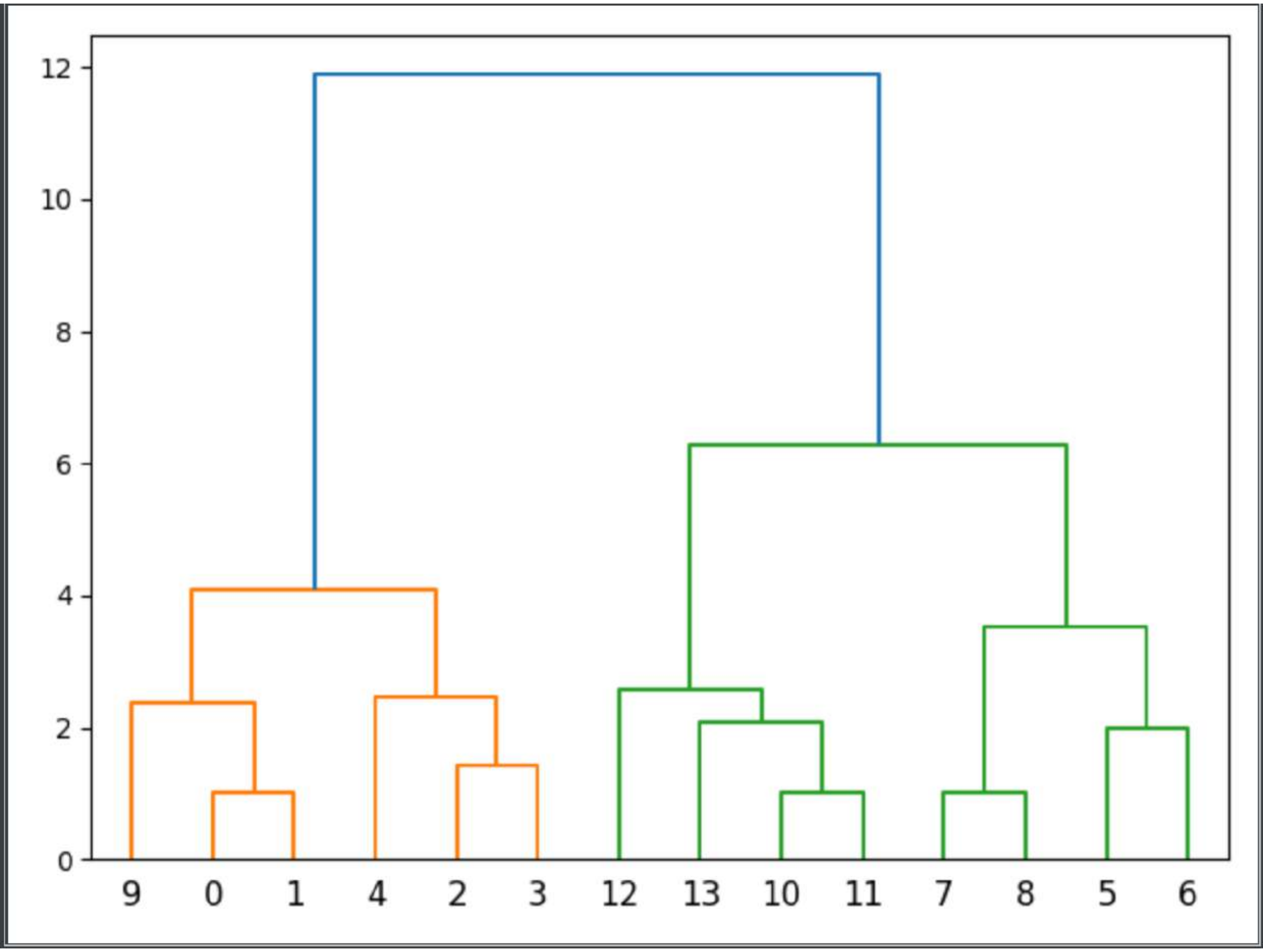
$$O_{i,j} = \#\{x = i \cap y = j\}$$

Finalement, il est possible de calculer l'écart relatif entre les deux variables :

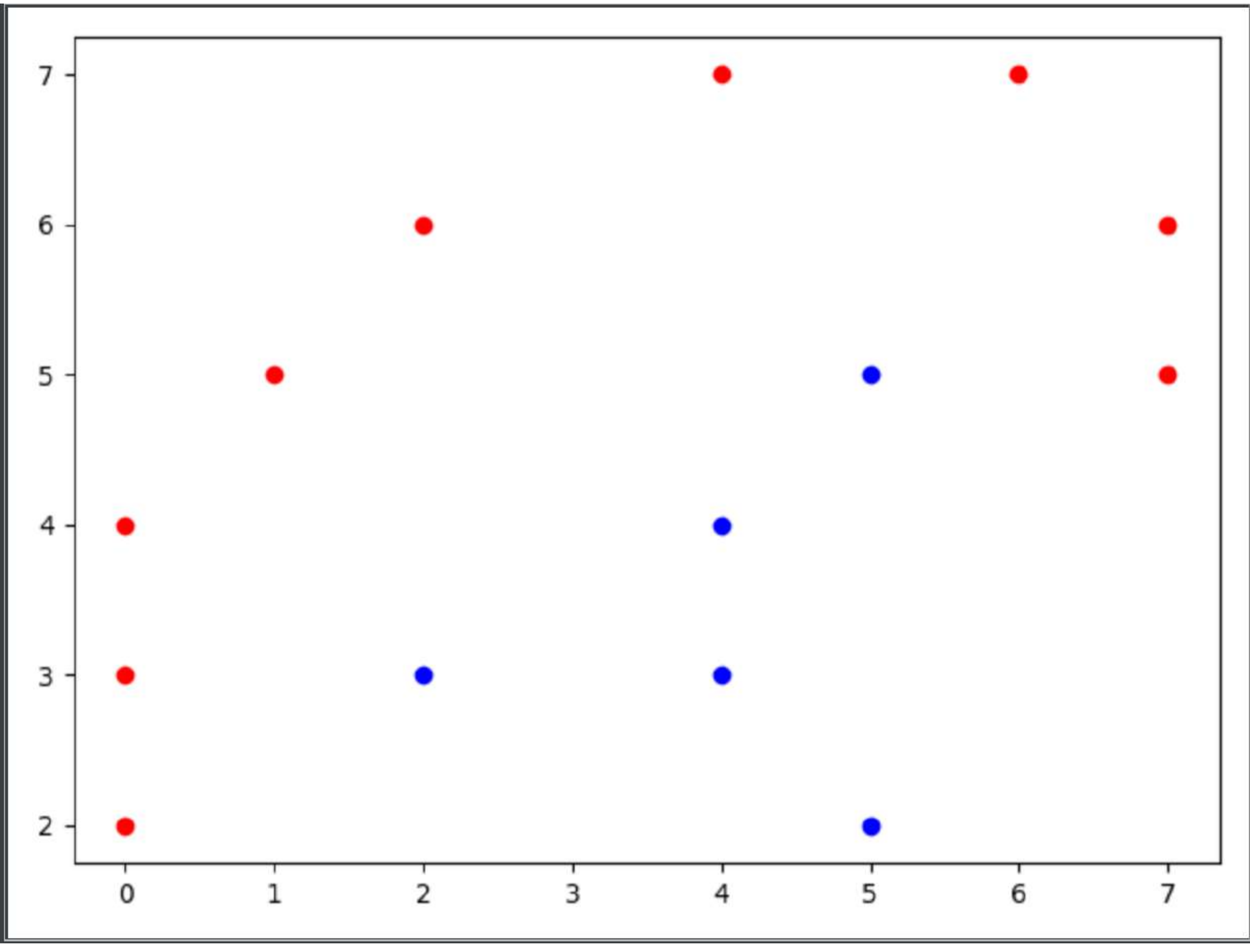
$$T = \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

T faible signifie que les variables sont indépendantes.

b) Pour trouver le nombre maximum de points pulvérisables, on peut utiliser la dimension de Vapnik-Chervonenkis.



12
10
8
6
4
2
0
9 0 1 4 2 3 12 13 10 11 7 8 5 6


7
6
5
4
3
2
0
1
2
3
4
5
6
7