Exercice 1 :

a)

Matrice de confusion Y_Est_1 :

		Actual T	Actual F
Predicted	T	6	8
	F	1	2

et Y_Est_2 :

	T	F
T	4	1
F	3	9

Risque empirique Est_1 :

$$D(\hat{f}) = 8 + x$$

Risque empirique Est_2 :

$$D(\hat{f}) = 1 + 3x$$

- Avantage Est_1 :

$$8 + x < 1 + 3x \iff 7 < 2x$$

$$\iff \boxed{x > \frac{7}{2}}$$

- Avantage Est_2 :

$$8 + x > 1 + 3x \iff 2x < 7$$

$$\iff \boxed{x < \frac{7}{2}}$$

b)

Exercice 1 : (suite)

d)

$Y=T$	1	2
X_1	5	2
X_2	2	5

$Y=F$	1	2
X_1	3	7
X_2	6	4

e) $Y=T$

$$P(Y=T \mid X_1=1, X_2=1) = P(Y=T) \cdot P(X_1=1 \mid Y=T) \cdot P(X_2=1 \mid Y=T)$$
$$= \frac{7}{17} \cdot \frac{5}{7} \cdot \frac{2}{7} \simeq 0{,}084$$

$$P(Y=T \mid X_1=1, X_2=2) = \frac{7}{17} \cdot \frac{5}{7} \cdot \frac{5}{7} \simeq 0{,}210$$

$$P(Y=T \mid X_1=2, X_2=1) = \frac{7}{17} \cdot \frac{2}{7} \cdot \frac{2}{7} \simeq 0{,}034$$

$$P(Y=T \mid X_1=2, X_2=2) = \frac{7}{17} \cdot \frac{2}{7} \cdot \frac{5}{7} \simeq 0{,}084$$

$Y=F$

$$P(Y=F \mid X_1=1, X_2=1) = P(Y=F) \cdot P(X_1=1 \mid Y=F) \cdot P(X_2=1 \mid Y=F)$$
$$= \frac{10}{17} \cdot \frac{3}{10} \cdot \frac{6}{10} \simeq 0{,}106$$

$$P(Y=F \mid X_1=1, X_2=2) = \frac{10}{17} \cdot \frac{3}{10} \cdot \frac{4}{10} \simeq 0{,}071$$

$$P(Y=F \mid X_1=2, X_2=1) = \frac{10}{17} \cdot \frac{7}{10} \cdot \frac{6}{10} \simeq 0{,}247$$

$$P(Y=F \mid X_1=2, X_2=2) = \frac{10}{17} \cdot \frac{7}{10} \cdot \frac{4}{10} \simeq 0{,}165$$

$(X_1=1, X_2=1) \rightarrow Y=F$

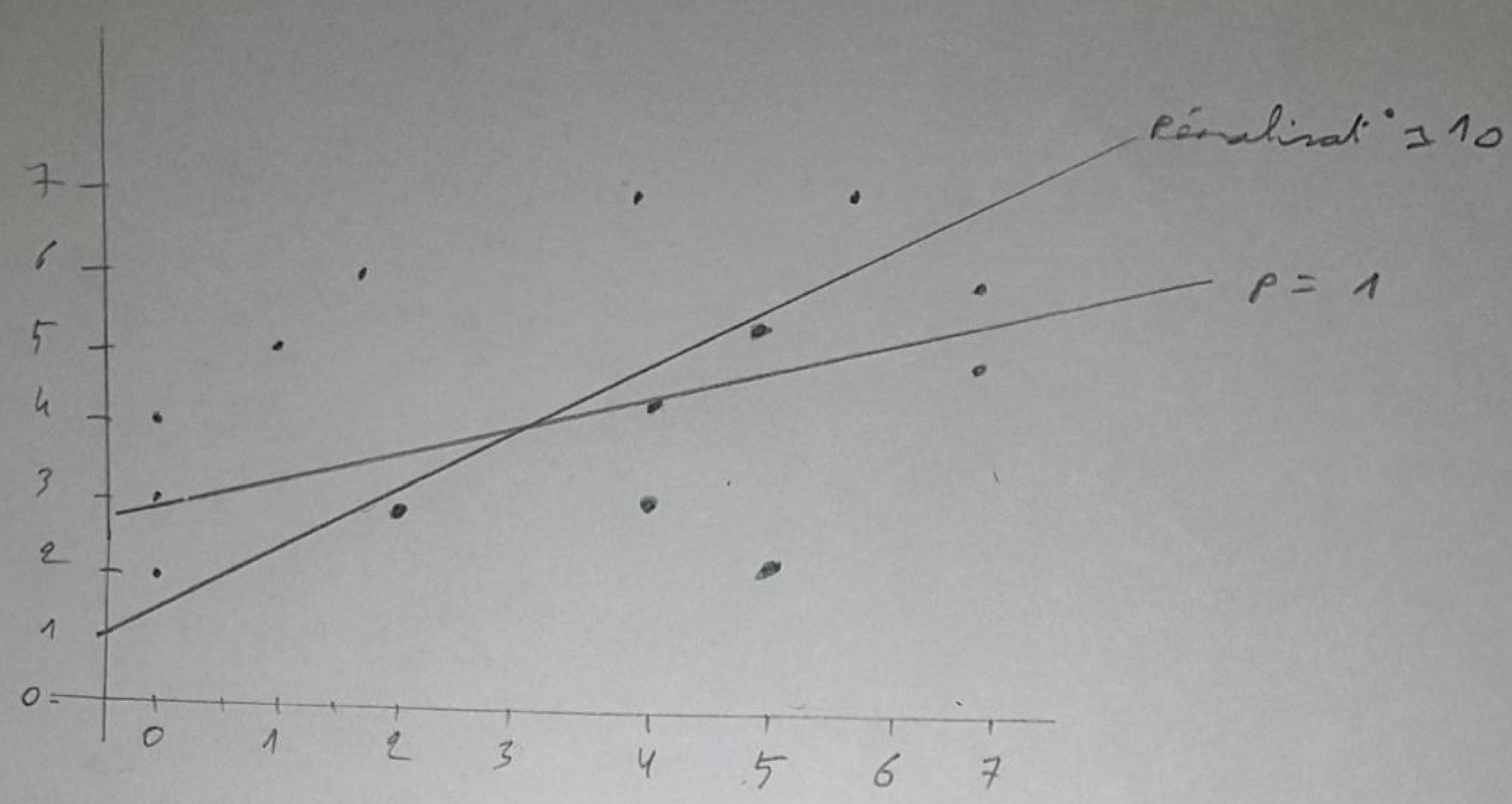
$(X_1=1, X_2=2) \rightarrow Y=T$

$(X_1=2, X_2=1) \rightarrow Y=F$

$(X_1=2, X_2=2) \rightarrow Y=F$

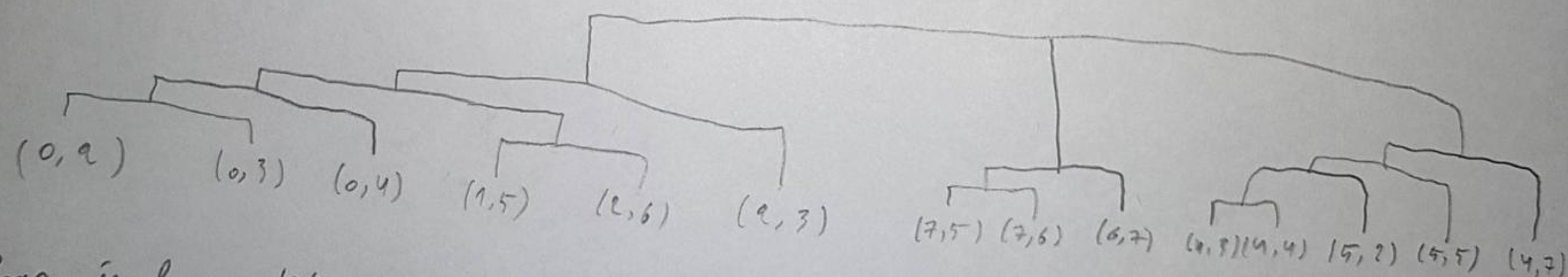
Exercice 2 :

a)



- Le nombre de classe étant de 2, on peut utiliser un k-means ou bien de l'agglomerative clustering ou encore un knn.

b)



- Pour évaluer (~~...~~) le coût d'intégrat°, on prend une fonction

c) Il n'est pas possible de les séparer avec un SVM linéaire sans commettre aucune erreur de prédict°.

d) cf schéma.

e) On peut plonger les données dans un nouvelle espace en utilisant un cercle. $x_3 = f(x_1, x_2) = x_1^2 + x_2^2$

où une sphère. $x_3 = f(x_1, x_2) = \left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2$.

Exercice 3 :

a) Pour savoir si deux variables sont indépendantes, j'effectuerais un test du chi-2,
Ce test reste tout de même pas parfait, et on doit fixer le taux de risque d'erreur (5% choisi par défaut).
Ce test va alors nous donner un p-value et si elle est endessous de 0.05 Alors les variables sont dépendantes et si la p-value > 0.05, Alors les deux variables sont Indépendantes.
Rq : (Dépendante ⇒ Rejet Hypothèse nulle et Indépendantes ⇒ Accept H0).

b) Pour trouver le nombre maximum de points pulvérisables à la surface d'une sphère, je commencerais par essayer de pulvériser 1, 2, 3, 4, ... points et ainsi de suite.
J'essayerai aussi de prendre des formes géométriques plus simplifié tel que le cercle, et si on atteint un maximum pour le cercle, on en déduis que ça l'est empiriquement aussi pour la sphère.

c) risque : À quel point notre algorithme est bon.

Ambiguité : À quel point il se trompe.