

FTAL

Exercice 1

a) a calcule les matrices de confusioa.

EST 1	T	F
Y=T	6	1
Y=F	8	2

EST 2	T	F
Y=T	4	3
Y=F	1	9

Le risque empirique pour EST 1 est donc :

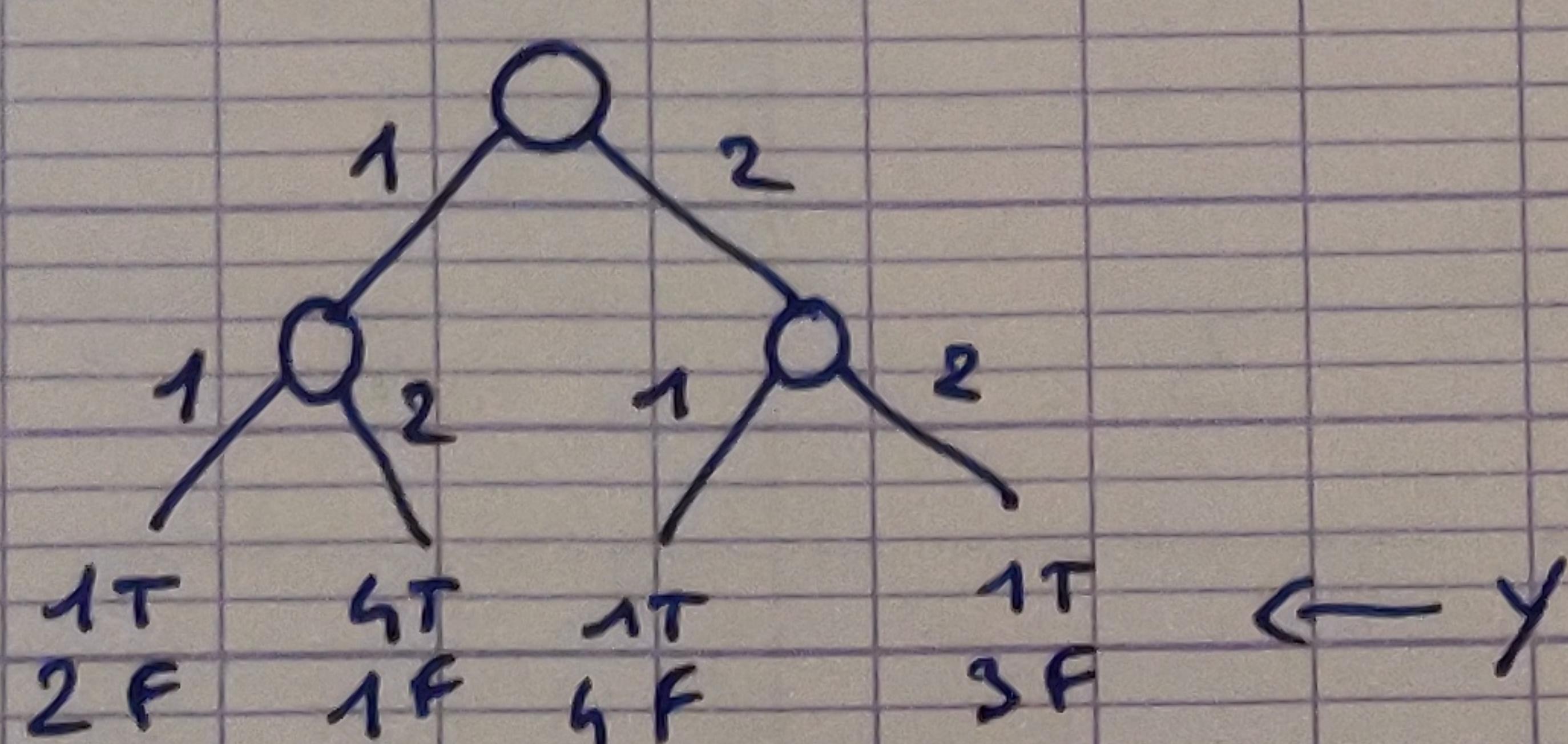
$$\frac{8}{10} + \alpha$$

Pour EST 2: $\frac{1+3\alpha}{10}$

Le risque de EST 1 est donc inférieur à EST 2
 si $\alpha < \frac{7}{2}$

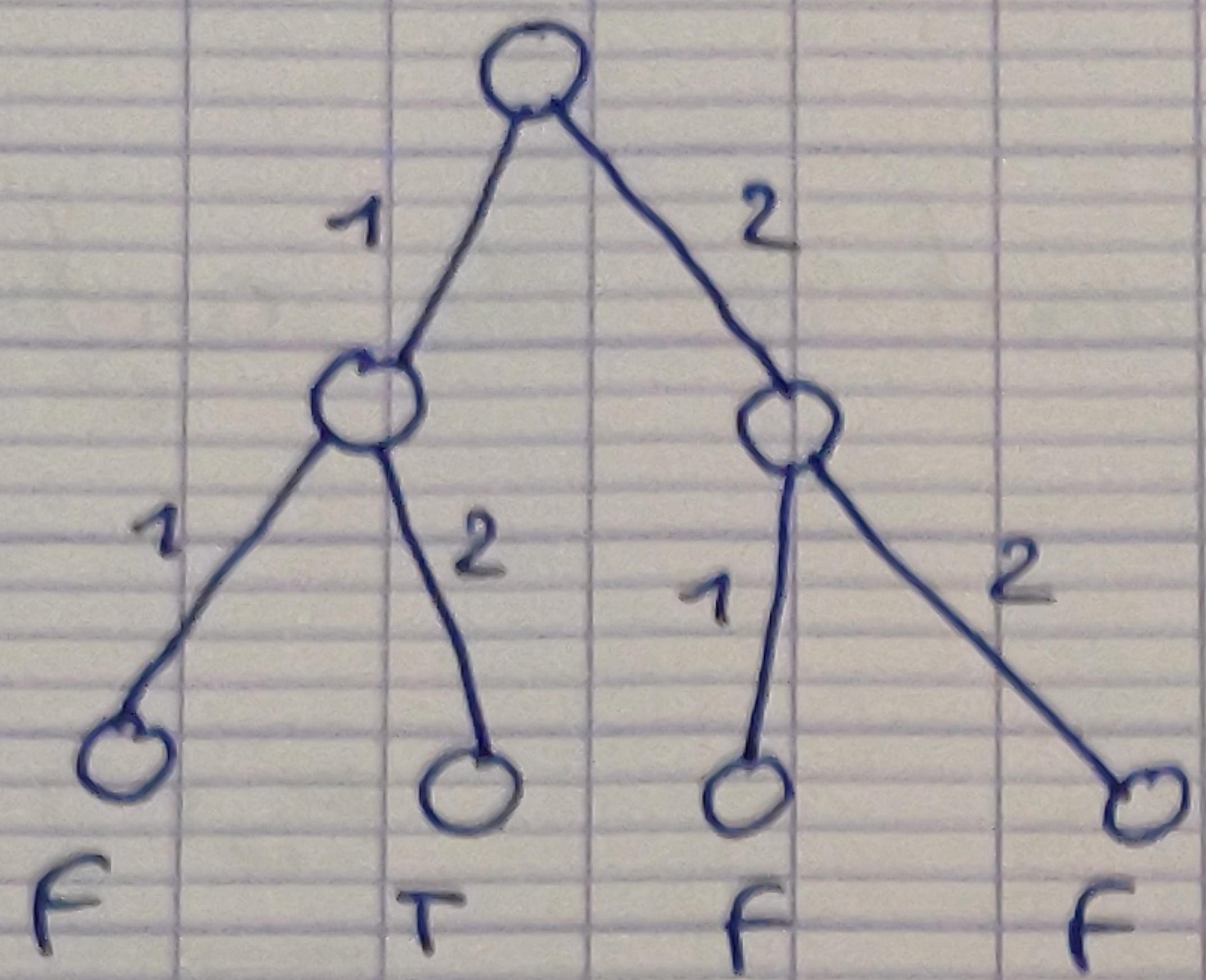
b)

On peut construire un arbre dépendant de x_1 et x_2



Sachant qu'on prendra deux fois plus un faux négatif.

On en déduira que l'estimateur \hat{d}_{id} est :



c)

d)

$Y = \text{True}$	$x = 1$	$x = 2$
x_1	5	2
x_2	2	5

$Y = \text{False}$	$x = 1$	$x = 2$
x_1	3	7
x_2	6	4

e)

On calcule $P(Y=T | (1,1))$ vs $P(Y=F | (1,1))$

$$\begin{aligned} P(Y=T | (1,1)) &= P(Y=T) \times P(X_1=1 | Y=T) \times P(X_2=1 | Y=T) \\ &= \frac{7}{17} \times \frac{5}{7} \times \frac{2}{7} \\ &= \frac{10}{119} \end{aligned}$$

(on ignore le dénominateur de la fraction de base car on va les comparer avec la même chose pour $Y=F$, qui a le même dénominateur)

$$\begin{aligned} P(Y=F | (1,1)) &= P(Y=F) \times P(X_1=1 | Y=F) \times P(X_2=1 | Y=F) \\ &= \frac{10}{17} \times \frac{3}{10} \times \frac{6}{10} \\ &= \frac{9}{85} \end{aligned}$$

donc $P(Y=F | (1,1)) > P(Y=T | (1,1))$

Pour le couple $(1,1)$ on choisit F

De la même manière, $P(Y=T | (1,2)) > P(Y=F | (1,2))$

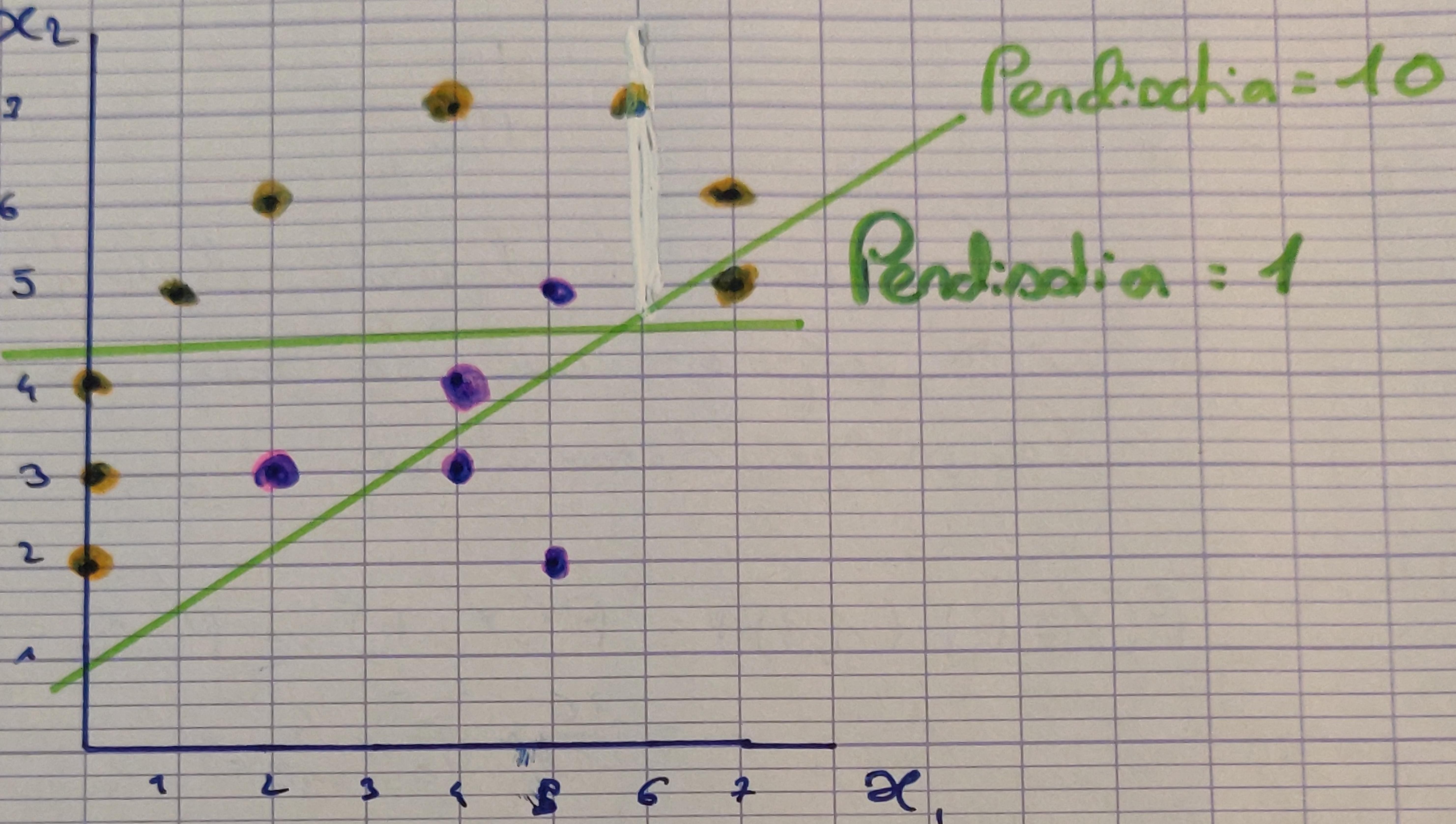
donc pour le couple $(1,2)$ on choisit T

De la même manière pour les couples $(2,1)$ et $(2,2)$ on a F et F

on a donc FT FF. On obtient le même estimateur que celui produit plus haut.

Exercice 2 :

a)



Les méthodes non supervisées à même de les détecter sont :

Le clustering spectral

Le clustering agglomératif (avec single linkage)

DBSCAN

OPTICS

b)

c)

Non il n'est pas possible de les séparer avec un
surm linéaire, les deux classes sont inseparables avec une
droite de manière parfaite.

On voit graphique.

e)

Exercice 3:

a) On vérifie le ratio de présence d'une classe
dans une autre. Si les deux classes sont indépendantes,
ce ratio devrait tourner autour de 50%. Si ce ratio
tend vers 100%, les données sont fortement corrélées et donc dépendantes,
et si tend vers 0, négativement corrélées et donc indépendantes

b) on doit commencer par choisir le séparateur, un hyperplan
par exemple. Ensuite, on ajoute des points 1 à 1 et on
regarde si on arrive à les séparer dans cette configuration
de deux classes. quand on ne peut plus les séparer on a atteint le nombre
maximal.

c)

Le risque caractérise la connaissance des possibilités et cherche
à quantifier leur probabilité d'occurrence.
L'ambiguïté caractérise la non connaissance des possibilités pouvant
survenir, donc in calculables.