FTML

Victor Chemla

Exercice 1 :

a) EST_1

On a la fonction de perte qui donne 1 si
$Y = F$ et $Y\_EST = T$ et
X si $Y = T$ et $Y\_EST = F$.

Comme on ne pénalise pas une bonne prédiction on aura 0.

On a $\Sigma(\tilde{\ell}(Y\_EST\_1)) = 8 + X$

On fait de même pour EST_2

$$\Sigma(\tilde{\ell}(Y - EST\_2)) = 3X + 1$$

On voit ici que l'estimateur 2 est mieux tant que $X < 2$ sinon l'estimateur 1 sera meilleur.

b) L'estimateur qui minimise le risque empirique est donc l'estimateur ou $\tilde{\ell} = Y$.

c) Ici nous ne sommes pas obligé de nous placer dans le cadre bayésien naïf car :

- Tout d'abord on voit que plusieurs réponses sont possibles pour la même entrée. On est ici sur un cas de surjection. Donc non-déterministe.

  Ex : à la ligne 1 et 4

- De plus on a trop de combinaison comparé au nombre de possibilité des variables

d)

X_1	X_2	$Y = T$	$Y = F$
1	1	1	2
1	2	4	1
2	1	1	4
2	2	1	3

e) $$P(Y=T) = \sum_{i=0}^{x_1} \sum_{j=0}^{x_2} \frac{7}{17} * \frac{x_1[i]}{7} * \frac{x_2[j]}{7}$$

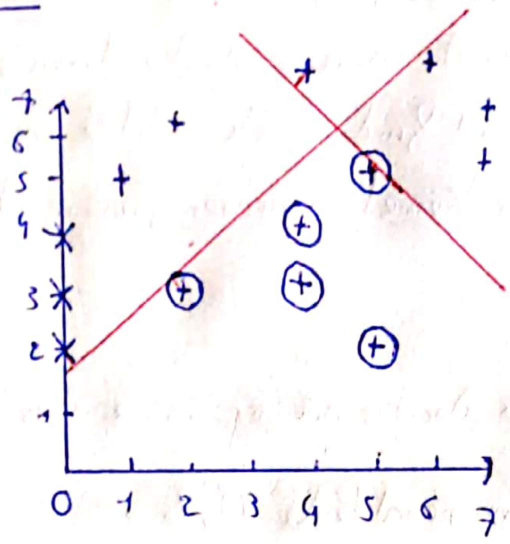
$$= 0{,}41$$

$$P(Y=F) = \sum_{i=0}^{x_1} \sum_{j=0}^{x_2} \frac{10}{17} * \frac{x_1[i]}{10} * \frac{x_2[j]}{10}$$

$$= 0{,}59$$

Exercice 2 :

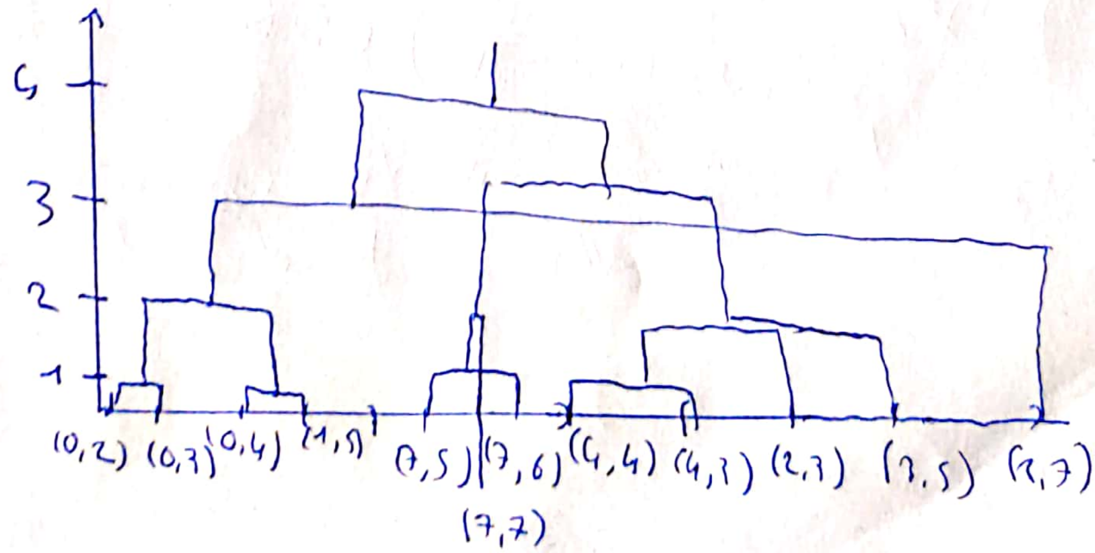
a)



Afin de distinguer c'est deux classes de manière non-supervisé nous pouvons utilisé :

- knn : Nearest Neighbours. Nous prenons le plus proche voisin à chaque fois.
- CAH : classification ascendante hiérarchique

b)



La fonction d'intégration est la distance donc $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

c) Ici il est impossible de séparer parfaitement avec un SVM linéaire. En effet, il n'existe aucune droite linéaire qui arrive à couper notre droite. Il faudrait alors utilisé un Kernel pour y arriver (peut-être un Kernel Gaussien pourrait être une bonne idée).

d) Vous retrouverez les deux droites rouges issue des SVM linéaires.

On peut estimer pour les deux une pénalité de 0,2.

e) Ici nous proposons un Kernel Gaussien. Afin de pouvoir avoir le paquet de S plus proche de la pointe Gaussien :

On aura donc $x_3 = e\left(-\frac{d(x_i, x_j)^2}{2\sigma}\right)$

ou x_i et x_j sont respectivement x_1 et x_2

Exercice 3 :

a) Si nous avons un tableau croisé d'effectifs. Il nous suffit de faire le test du χ^2 (khi-deux) :

Tel que : $T = \left(\sum_{i,j} \frac{n_{ij}^2}{n_{i\cdot} n_{\cdot j}} - 1 \right)$

Avec $n_i = \sum_0^i n_i$

Si T est suffisamment petit alors les valeurs sont liées sinon elles sont indépendantes.

b) Il y a une infinité de pulvérisation sur une sphère car on pourra toujours faire un demi-plan

c) Le risque c'est la probabilité de faire une erreur alors que l'ambiguïté c'est plus la probabilité de répondre deux résultats différents pour une même entrée