SCIENCES SORBONNE UNIVERSITÉ

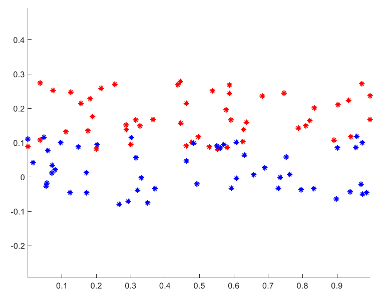
Examen de machine learning
Année universitaire 2020-2021

Nom:	
Prénom:	
Parcours	

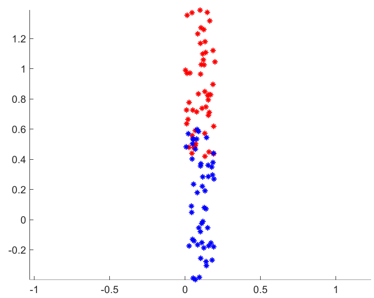
Exercice 1 – 6 pts

Toutes les questions sont indépendantes

On considère les deux ensembles de données suivants :



Données 1



Données 2

1. **(1 pt)** Pour l'ensemble de données 1, quel axe (une fois les données projetées sur cet axe) a le plus grand pouvoir discriminant ? Unifiant ?

> Plus grand pouvoir discriminant : axe vertical
> Plus grand pouvoir unifiant : axe vertical

2. **(1 pt)** Même question pour l'ensemble de données 2

> Plus grand pouvoir discriminant : axe vertical
> Plus grand pouvoir unifiant : axe horizontal

3. **(1 pt)** Considérons l'ensemble de données 1, tracer sur la figure correspondant le premier axe de projection obtenu par ACP.

> C'est l'axe horizontal

4. **(1.5 pts)** Les matrices de covariance des deux ensembles de données ont été estimées :

$$A = \begin{pmatrix} 0.0034 & 0.0074 \\ 0.0074 & 0.1939 \end{pmatrix} \text{ et } B = \begin{pmatrix} 0.0910 & 0.0003 \\ 0.0003 & 0.0093 \end{pmatrix}$$

Attribuer chaque matrice de covariance à l'ensemble de données correspondant.

A : données 2
B : données 1

5. **(1.5 pts)** Pour l'ensemble de données 2, le premier axe de projection estimé par l'ACP est défini par le vecteur $(0.0390 \quad 0.9992)^T$. Quel sera la nouvelle coordonnée du point $(0.1 \quad 1)^T$, une fois projeté sur cet axe ?

$(0.0390 \quad 0.9992)^T.(0.1 \quad 1)$=1.0031

Exercice 2 – 3 pts

Toutes les questions sont indépendantes

On considère un problème de classification à 3 classes C1, C2, C3. Le jeu de données est constitué de seulement 20 exemples. On décide donc d'estimer les résultats en utilisant une K-fold validation avec K=4.

Les matrices de confusion obtenues pour chaque ensemble de validation sont :

$$\begin{pmatrix}2 & 1 & 0\\0 & 1 & 0\\0 & 1 & 0\end{pmatrix} \qquad \begin{pmatrix}0 & 0 & 0\\0 & 2 & 1\\0 & 0 & 2\end{pmatrix} \qquad \begin{pmatrix}1 & 0 & 1\\0 & 2 & 1\\0 & 0 & 0\end{pmatrix} \qquad \begin{pmatrix}3 & 1 & 0\\0 & 0 & 0\\0 & 0 & 1\end{pmatrix}$$

1. **(1.5 pts)** Déterminer le nombre d'exemples par classe de la base de données

C1 : 9 - C2 : 7 - C3 : 4 (il suffit de sommer les lignes de toutes les matrices. On retrouve bien que la base est composée de 20 exemples.

2. **(1.5 pts)** Quel est le taux de reconnaissance de ce classifieur ?

T1 = 3/5, T2 = 4/5, T3 = 3/5, T4 = 4/5
Taux = 5T1+T2+T3+T4)/4 = 14/20 = 0.65

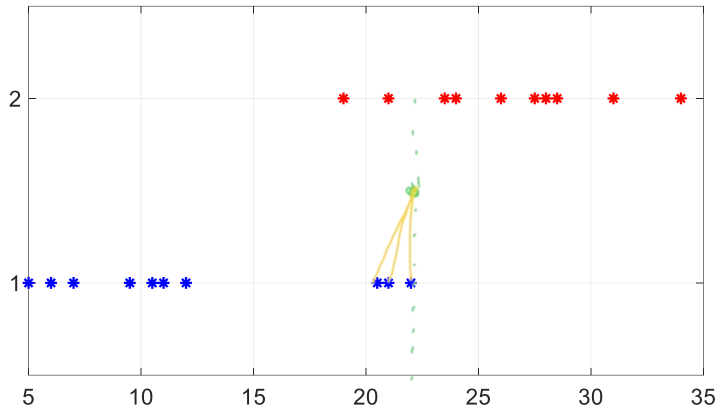
Exercice 3 – 8 points

Toutes les questions sont indépendantes

On considère un problème de classification de dimension 1, à deux classes. La classe 1 est deux fois plus probable que la classe 2 (P(C1)=2/3 et P(C2)=1/3). Chaque classe est définie par un ensemble de données de 10 points :

C1	5	10.5	21	20.5	9.5	12	7	6	22	11
C2	21	34	27.5	23.5	26	28	19	31	28.5	24

Les classes sont également représentées ci-dessous :



On souhaite classer un nouvel exemple tel que $x = 22$. On décide de résoudre ce problème de classification en utilisant la classification bayésienne, ce qui demande d'estimer les vraisemblances de chaque classe : $p(x/C1)$ et $p(x/C2)$.

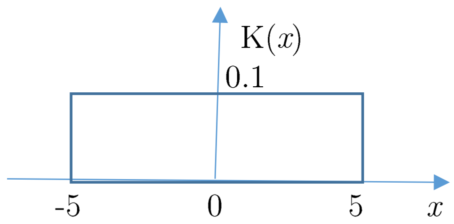
1. **(2 pts)** On décide dans un premier temps de modéliser ces vraisemblances par un histogramme normalisé. Les cases de ces histogrammes débutent à 0 et ont une largeur de 10. Donner la valeur de $p(x/C1)$ et $p(x/C2)$ pour $x = 22$. En déduire les valeurs de $p(C1/x)$ et $p(C2/x)$. Conclusion ?

$p(x/C1) = 0.3$ et $p(x/C2) = 0.7$

$p(C1/x) = \frac{0.3*2/3}{0.3*2/3+0.7*1/3}$ et $p(C2/x)=\frac{0.7*1/3}{0.3*2/3+0.7*1/3}$

⇨ L'exemple sera classé C2

2. **(2 pts)** On estime ensuite ces vraisemblances de manière plus robuste, à l'aide des fenêtres de Parzen. On utilise pour cela le noyau :



Donner la valeur de $p(x/C1)$ et $p(x/C2)$ pour $x = 22$. En déduire les valeurs de $p(C1/x)$ et $p(C2/x)$. Conclusion ?

$p(x/C1)$= 0.03 (point en 21, 20.5 et 22)
$p(x/C2)$= 0.05 (point en 21, 23.5, 26, 19et 24)

$p(C1/x) = \frac{0.03*2/3}{0.0.03*2/3+0.05*1/3}$ et $p(C2/x) = \frac{0.05*1/3}{0.03*2/3+0.05*1/3}$

⇨ L'exemple sera classé C1

3. **(2 pts)** On décide maintenant d'estimer ces vraisemblances de manière paramétrique en utilisant une loi normale de paramètre (μ, σ). Retrouver l'expression analytique de μ avec la règle du maximum de vraisemblance

VOIT cours

4. **(2 pts)** On souhaite maintenant tester une approche discriminative pour classer l'exemple $x = 22$. Donner le résultat de classification en utilisant le 1PPV, puis le 3PPV, en justifiant à chaque fois votre réponse.

1PPV : l'exemple le plus proche est 22 (C1) ➔ classe 1
3PPV : 22 (C1), 21 (C1) et 21 (C2)➔ classe 1

Exercice 4 – 3 pts

On souhaite détecter des piétons (positifs) dans les images. La base est constituée de 1000 exemples, dont 200 exemples positifs.

Pour la détection, nous disposons d'un algorithme qui dépend d'un paramètre s. En utilisant 3 valeurs de ce paramètre, nous obtenons :

	Nb piétons détectés piétons	Nb NON piétons détectés piétons
$s1$	0	0
$s2$	100	400
$s3$	200	600

Pour chaque valeur de s, déterminer la sensibilité et la spécificité

Il y a 200 exemples positifs et 800 négatifs.
Le nombre de piétons détectés piétons correspond à TP et le Nb de non piétons détectés piétons à FP
On a donc les matrices de confusion :

$\begin{bmatrix} 0 & 200 \\ 0 & 800 \end{bmatrix}$ sens=0 - spe = 1	$\begin{bmatrix} 100 & 100 \\ 400 & 400 \end{bmatrix}$ sens=0.5 - spe = 0.5	$\begin{bmatrix} 200 & 0 \\ 600 & 200 \end{bmatrix}$ sens=1 - spe = 0.25