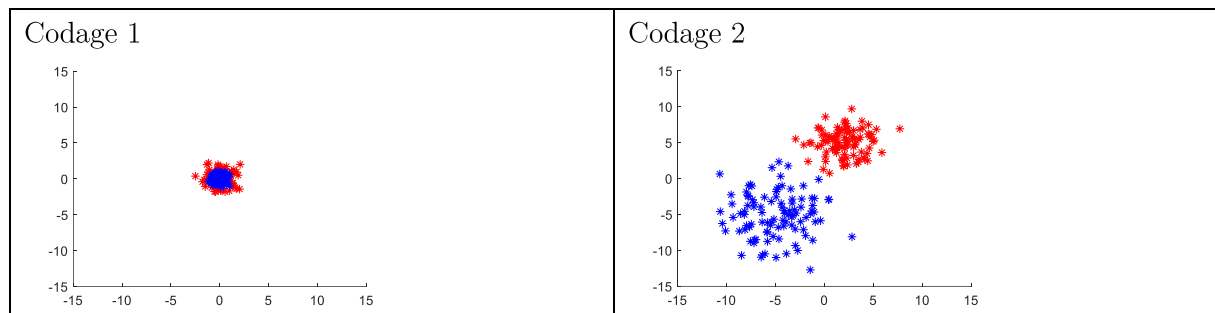


Nom:	
Prénom:	

Exercice 1 – 4 pts

Pour un même problème de classification à deux classes, on dispose de deux codages, chacun de dimension 2. Les 200 exemples de la base sont représentés sur les deux figures ci-dessous avec chacun de ces codages.



Quel codage a le plus fort pouvoir discriminant ? Pourquoi ?

Quel codage a le plus fort pouvoir unifiant ? Pourquoi ?

Pour le codage 2, les deux vecteurs propres de la matrice de covariance sont $\mathbf{v}_1 = \begin{bmatrix} -0.8 \\ 0.6 \end{bmatrix}$ et $\mathbf{v}_2 = \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}$. Lequel correspond à l'axe de projection obtenu par ACP ?

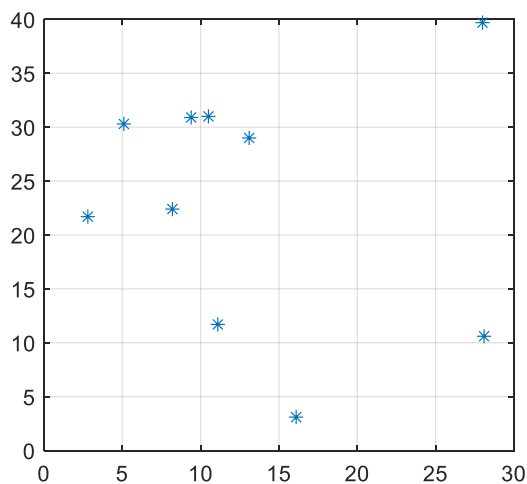
On utilise cet axe pour réduire la dimension des données et donc passer en dimension 1. Donner la valeur du nouveau codage \mathbf{x}_m (de dimension 1) obtenu pour l'exemple $\mathbf{x} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$ en posant vos calculs.

On souhaite maintenant reconstruire l'exemple \mathbf{x} à partir de \mathbf{x}_m . Donner les valeurs du point reconstruit \mathbf{x}_r (de dimension 2) dans l'espace d'origine.

Exercice 2 – 4 pts

On dispose de 10 échantillons \mathbf{x} , de dimension 2, dont on souhaite estimer la densité de probabilité $\hat{p}(\mathbf{x})$. Ces échantillons sont donnés et représentés ci-après.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
5.1	8.2	10.5	9.4	16.1	11.1	28.1	13.1	28	2.8
30.3	22.4	31.0	30.9	3.1	11.7	10.6	29	39.7	21.7



Dans un premier temps, on estime la densité de probabilité $\hat{p}(\mathbf{x})$ en utilisant un histogramme 2D dont les cases correspondent à celles de la figure. Que vaudront $\hat{p}\left(\begin{smallmatrix} 9 \\ 29 \end{smallmatrix}\right)$ et $\hat{p}\left(\begin{smallmatrix} 10.5 \\ 24 \end{smallmatrix}\right)$?

Dans un second temps, on utilise les fenêtres de Parzen et le noyau 2D suivant :

$$K(x, y) = \begin{cases} A & \text{si } |x| < 2.5 \text{ et } |y| < 2.5 \\ 0 & \text{sinon} \end{cases}$$

Déterminer la valeur de A, puis celles $\hat{p}\left(\begin{smallmatrix} 9 \\ 29 \end{smallmatrix}\right)$ et $\hat{p}\left(\begin{smallmatrix} 10.5 \\ 24 \end{smallmatrix}\right)$, en fonction de A.

Exercice 3 – 5 pts

On considère un HMM à 3 états : s_1 =soleil, s_2 = pluie, s_3 =nuage. Les observations correspondant à ces états sont directement l'état du ciel. On a donc, pour le premier état $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, pour le second, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ et pour le troisième $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. A l'instant initial, chaque état est équiprobable. Pour les instants suivants, on a 2 fois plus de chance de rester dans le même état que de changer d'état et si on change d'état, les deux autres états sont équiprobables.

Déterminer la matrice de transition A, le vecteur de probabilité initiale Π et la matrice d'observation B.

Déterminer la probabilité de la séquence y = « soleil, soleil, nuage, soleil, pluie » en utilisant l'algorithme forward.

Exercice 4 – 3 pts

On dispose maintenant d'une base de données de 100 exemples. Afin d'estimer les performances de classifieurs de manière fiable, on adopte une approche 4-folds validation.

Rappeler en quoi consiste la 4-folds validation

On dispose de deux classifieurs pour ce problème que l'on veut comparer. Les résultats pour chaque fold sont donnés par :

Classifieur1	75	90	80	85
Classifieur1	75	73	75	74

Déterminer le taux de reconnaissance final obtenu par chaque classifieur.

Quel classifieur a le plus fort biais ? Quel classifieur a la plus forte variance ? Pourquoi ?

Exercice 5 – 4 pts

Considérons un problème de détection de visages dans les images. Un détecteur a été testé sur une base de test comportant 100 images de visage (positifs) et 100 images de non visage (négatifs). En faisant varier le seuil de détection (4 seuils), le détecteur amène aux vrais positifs (TP) et faux positifs (FP) suivants :

	Seuil1	Seuil2	Seuil3	Seuil4
TP	20	40	60	80
FP	20	80	90	95

Tracer la courbe ROC correspondant à ce détecteur en précisant bien les échelles ainsi que les 4 points de fonctionnement correspondant aux 4 seuils

