| université de BORDEAUX | Année universitaire : **2019/2020**<br>**DS Analyse, Classification, Indexation des Données**<br>Code UE : **4TIN703U** | | |
|---|---|---|---|
| Collège<br>Sciences & Technologies | Date : **8 novembre**<br>**Documents non autorisés** | Heure : **8h00** | Durée : **1h30**<br>**4 pages** |

Nom : NAVARRO
Prénom : Robin
Groupe : 5

---

**Exercice 1 : Descente de gradient**

Pour rappel, la méthode de descente du gradient consiste à rechercher le minimum d'une fonction en se basant sur le signe de sa dérivée.

Soit une fonction $f(x)$ et $f'(x)$ sa dérivée par rapport à $x$. La technique de descente du gradient nous dit qu'un minimum de la fonction peut être trouvé en suivant l'algorithme suivant, étant donné un point de départ $x_{init}$ et un pas d'apprentissage $nu$. Le critère d'arrêt dépend d'une valeur $\epsilon$.

---

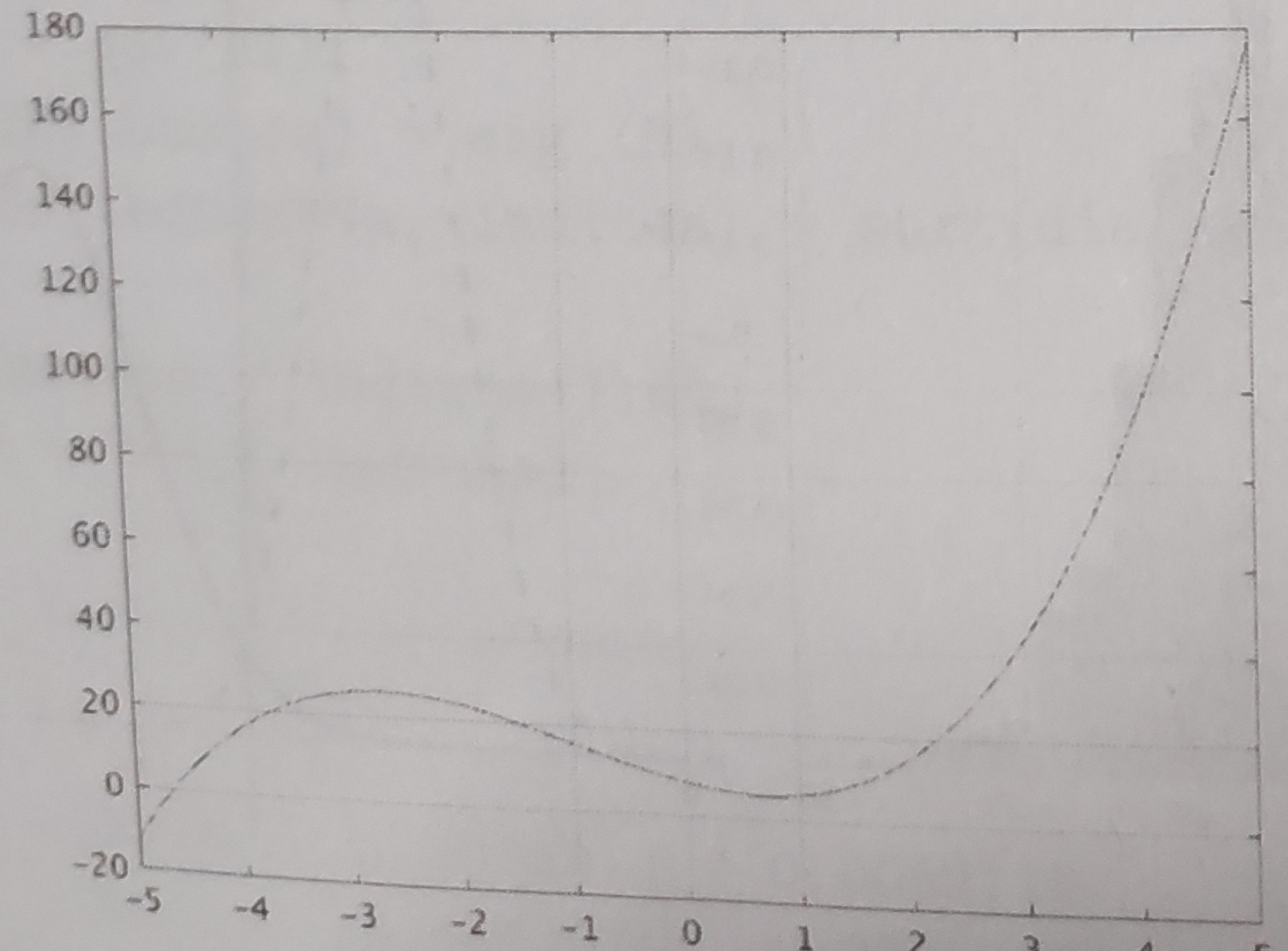
$x_{current} = x_{init} - nu \times f'(x_{init})$
distance=$|x_{current} - x_{init}|$

Tant que (distance $> \epsilon$)
  $x_{init} = x_{current}$
  $x_{current} = x_{init} - nu \times f'(x_{init})$
  distance = $|x_{current}\text{-}x_{init}|$

---

1. Expliquez l'algorithme en répondant aux questions suivantes :
   (a) Où intervient le signe de la dérivée ?
   (b) Quelle variable contient la position du minimum après exécution de l'algorithme ? Quelle est la valeur du minimum ?
   (c) Quelle est l'influence du paramètre $\epsilon$ ?
   (d) Pourquoi faut-il introduire un pas d'apprentissage ($nu$ dans l'algorithme) ? Que peut-il se passer si $nu$ est trop petit ? S'il est trop grand ?

2. On étudie la fonction $f(x) = x^3 + 3x^2 - 6x + 8$ dont la représentation graphique pour $-5 \leq x \leq 5$ est la suivante :




   (a) Calculez sa dérivée $f'(x)$.

(b) En prenant comme point de départ $x_{init} = -1$, comme pas d'apprentissage $nu = 0.1$, et en fixant $\epsilon = 0.1$, calculez les valeurs de $x_{current}$ jusqu'à l'arrêt de l'algorithme.

(c) Obtient-on le même résultat si on choisit $x_{init} = 4$ ? Si on choisit $x_{init} = -4$ ? Justifiez votre réponse sans faire de nouveaux calculs.

**Exercice 2 : Classifieur Bayésien (2 classes)**

1. Expliquez le principe des deux classifieurs vus en TD : classifieur bayésien basé sur le maximum de vraisemblance (ML) et classifieur bayésien basé sur le maximum a posteriori (MAP).
   Vous noterez :
   $p(C)$ la probabilité a priori de la classe $C$,
   $p(x|C)$ la probabilité qu'un individu ait $x$ pour descripteur sachant qu'il appartient à la classe $C$,
   $p(C|x)$ la probabilité qu'un individu soit de classe $C$ sachant qu'il a $x$ comme descripteur.

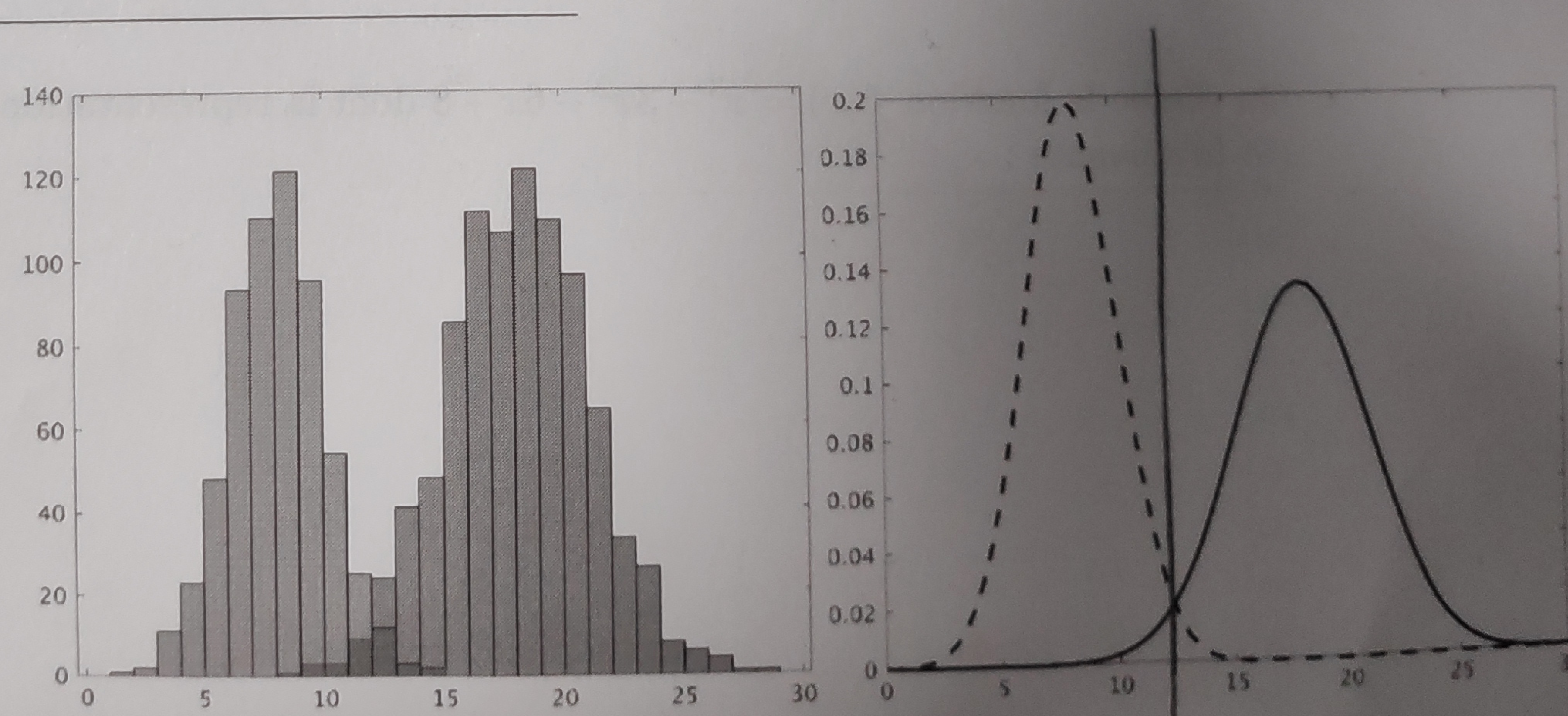
Dans la suite, on fait l'hypothèse que les distributions de probabilités sont gaussiennes.

2. Cas pratique A
   On fournit les résultats du code matlab suivant pour des données contenues dans les vecteurs $Train1$ et $Train2$. Le vecteur $Train1$ contient les exemples de la classe 1, le vecteur $Train2$ contient les exemples de la classe 2.

```
size1 = length(Train1)
size2 = length(Train2)
m1 = mean(Train1)
m2 = mean(Train2)
v1 = var(Train1)
v2 = var(Train2)
figure(1)
histogram(Train1);
hold on
histogram(Train2);
figure(2)
x = 0:0.5:30;
plot(x, normpdf(x, m1, sqrt(v1)), '-');
hold on
plot(x, normpdf(x, m2, sqrt(v2)), '--');
```

```
** CAS A **

size1 =   900
size2 =   600
m1 =   18.1149
m2 =   8.0862
v1 =   8.9536
v2 =   3.9938
```




(a) Donnez les probabilités des classes 1 et 2.

(b) Donner une valeur approximative du seuil $x_{seuil}$ de décision pour le classifieur ML (utilisez le sujet pour le construire géométriquement).

N° d

(c) Le seuil de décision pour le classifieur MAP sera-t-il supérieur ou inférieur à $x_{seuil}$ ? Expliquez.

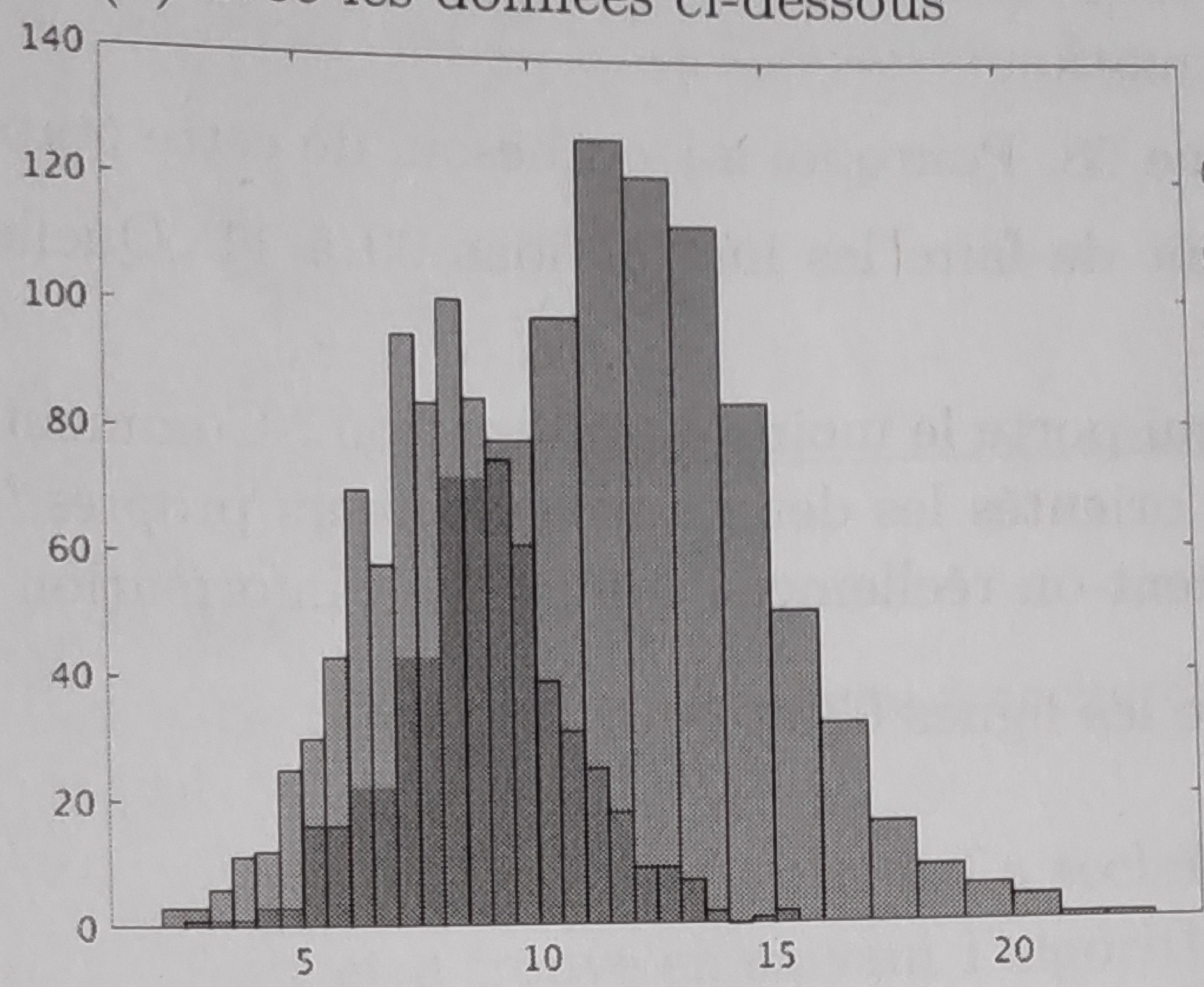
(d) Au vu des informations fournies, pouvez-vous prévoir la qualité des résultats des classifieurs ML et MAP entrainés avec ces données ? Expliquez.

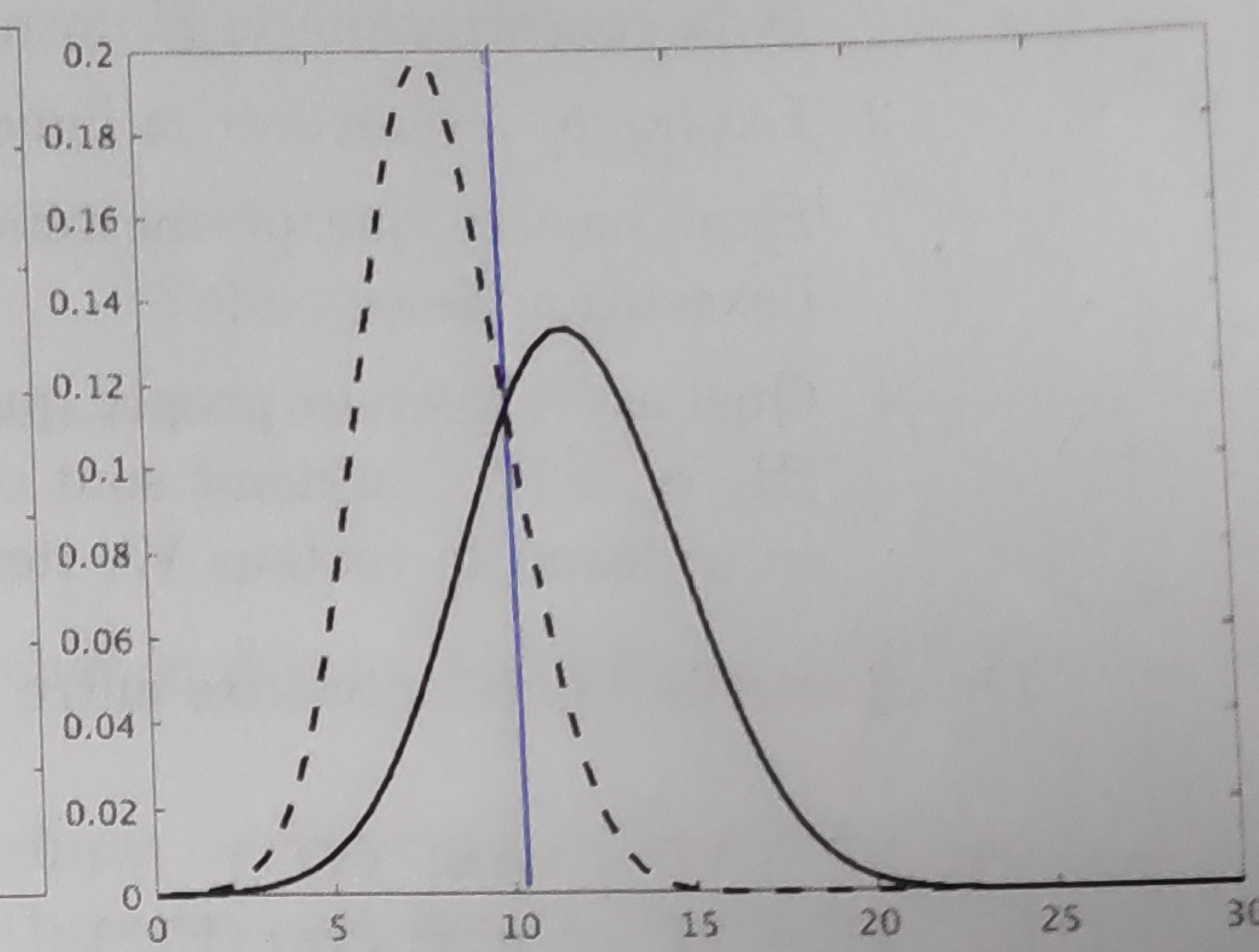
3. Cas pratique B

   Mêmes questions qu'au (2) avec les données ci-dessous

```
** CAS B **

size1 =   900
size2 =   900
m1 =   11.8488
m2 =   8.0958
v1 =   8.9255
v2 =   4.0485
```




## Exercice 3 : Réduction en dimension

La table suivante contient des données relatives à un échantillon de femmes et d'hommes. Il s'agit de mesures relatives à la taille en mètre de la personne, à son poids en kg et à la longueur de ses pieds en cm.

| S(ex) | H(eight) (m) | W(eight) (kg) | F(oot) size (cm) |
|---|---|---|---|
| M | 1.82 | 82 | 30 |
| M | 1.80 | 86 | 28 |
| M | 1.70 | 77 | 30 |
| M | 1.80 | 75 | 25 |
| F | 1.52 | 45 | 15 |
| F | 1.65 | 68 | 20 |
| F | 1.68 | 59 | 18 |
| F | 1.75 | 68 | 23 |

On veut réduire ces informations de façon à obtenir une seule valeur par personne. Pour cela on veut utiliser l'ACP. Soit le code Matlab suivant :

```
H = [ 1.82, 82, 30; 1.80, 86, 28; 1.70, 77, 30; 1.80, 75, 25];
F = [ 1.52, 45, 15; 1.65, 68, 20; 1.68, 59, 18; 1.75, 68, 23];
VT = [H; F] ;
M = cov(VT);
[VP, Lambda] = eig (M)
[SortedLambda, indices] = sort(diag(Lambda), 'descend');
d = 1;
indices = indices(1:d);
W = VP(:, indices);
```

Et la sortie du programme :

```
VP =
   0.9999628    0.0059912    0.0062125
  -0.0080014    0.3737038    0.9275136
   0.0032353   -0.9275287    0.3737378
```

```
Lambda =
   2.0734e-03              0             0
            0     3.8357e+00             0
            0              0    2.0330e+02
```

1. De manière générale dans l'ACP quelle est la relation qui existe entre les valeurs propres calculées et la conservation de l'information ?
2. Expliquez à quoi sert la ligne 08. Pourquoi a-t-on besoin de cette étape ?
3. Expliquez ce que permettent de faire les instructions 09 à 11. Quelle est la valeur de W après l'exécution de ce code ?
4. Quel est le vecteur propre qui porte le moins d'information ? Comment est-il orienté dans l'espace (H, W, F) ? Comment sont orientés les deux autres vecteurs propres ? Si on projette les données en utilisant le vecteur W, tient-on réellement compte de l'information de taille (H) ?

On ajoute les lignes suivantes entre les lignes 03 et 06 :

```
04. VT = VT - mean (VT);
05. VT = VT ./ sqrt(var(VT));
```

On obtient maintenant les résultats suivants :

```
VP =
  -0.31119  -0.76428   0.56484
   0.79718   0.11364   0.59295
  -0.51737   0.63479   0.57390

Lambda =
   0.051435          0          0
          0   0.204870          0
          0          0   2.743695
```

5. Expliquez à quoi servent les instructions ajoutées.
6. Quelle est maintenant la valeur de W ?
7. Pensez vous que cette nouvelle version permet de résoudre le problème pointé à la question 4 ? Justifiez.

FIN.

DE L'UE :
UVE DE : ACID
DE COURS :
NONYMAT : Groupe 5

d'intercalaire(s) employé(s)

Rclin
Étudiant : 2160 9361
Numéro de place
A RABATTRE

| NOTE | APPRÉCIATIONS EXPLIQUANT LA NOTE CHIFFRÉE : |
|---|---|
| 20/20 même un plus ! | Ex1 $\frac{8}{8}$ Ex2 : $\frac{6,5}{8}$ Ex3 : $\frac{6}{6}$ |

## Exercice 1 : Descente de gradient

1) a) Il intervient à la ligne $x_{current} = x_{init} - nu \times f'(x_{init})$. Il nous permet d'aller dans le sens de la pente.
Si il est négatif, on "avance", sinon, on "recule".

1

b) La variable $x_{current}$ contient la position du minimum à la fin de la fonction.

La valeur du minimum est alors $f(x_{current})$

1

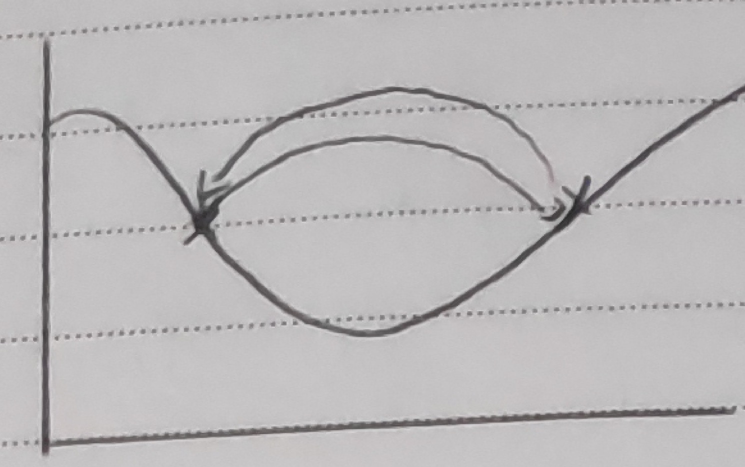
c) ~~Il influence la longueur des "pas" que nous allons effectuer sur la courbe.~~

d) Le paramètre c influe sur la précision du résultat. Plus il est petit, plus le resultat sera précis et inversement.

1

d) Il permet de changer la longueur des pas que nous allons effectuer.
Si il est ~~pas~~ trop grand, on risque

de faire des aller-retours.



S'il est trop petit, on risque de mettre trop de temps à atteindre le minimum. ou de s'arrêter trop tôt. 1

2) a) $f'(x) = 3x^2 + 6x - 6$ 1

b) $x$ courant : $-\frac{1}{10} \to 0{,}557 \to 0{,}7297253$

$\downarrow$

$0{,}7277310184$

c) Si on choisi 4, on aura environ le même résultat. On arrive de "l'autre côté" du minimum.

Avec $-4$, nous n'avons pas le même résultat. En effet, nous allons suivre la pente, et donc on va aller vers la gauche de la courbe. 1

## Exercice n°2 : Classifieur Bayésien

1) Avec le maximum de vraisemblance, on regarde la probabilité d'obtenir le feature $x$ sachant que l'individu est d'une classe $\omega$.

Donc, si $P(f=x \mid \omega_1) > P(f=x \mid \omega_2)$
alors $\omega_1$
sinon $\omega_2$

3

Avec le maximum à posteriori, on regarde la probabilité d'avoir une classe $\omega$, sachant que le feature est $x$.

Donc, si $P(\omega_1 \mid f=x) > P(\omega_2 \mid f=x)$
alors $\omega_1$
sinon $\omega_2$

2) a) $P(1)$ = size 1 / (size 1 + size 2) = 0,6

0,5 $P(2) = 1 - P(1) = 0,4$

b) $x$ seuil est environ à 12,5.

0,5

c) Pour MAP, on a :

$$P(\omega_1 | f = x) >? P(\omega_2 | f = x)$$

Mais nous avons la formule de Bayes :

$$P(A|B) = \frac{P(B|A) \times P(B)}{P(A)}$$

Notre classifieur MAP devient :

$$\frac{P(f=x | \omega_1) \; P(\omega_1)}{P(f=x)} >? \frac{P(f=x|\omega_2)P(\omega_2)}{P(f=x)}$$

Comme $P(f=x)$ est pareil des deux côté :

0,5 $$P(f=x | \omega_1) P(\omega_1) >? P(f=x | \omega_2) P(\omega_2)$$

On a donc le seuil de ML, mais déplacé avec les valeurs de $P(\omega_1)$ et $P(\omega_2)$.

Raisonnement correct et conclusion fausse...

Comme $P(\omega_1) > P(\omega_2)$, on aura un $x$ seuil supérieur.

d)

3) a) $P(1) = P(2) = 0,5$

0,5 b) Le seuil est environ à 10

0,5 c) Même explication que 2) c). Mais comme $P(1) == P(2)$, le seuil reste le même, $\simeq 10$

1

d) L'erreur est $\min(P(f=x|\omega_1), P(f=x|\omega_2))$

On peut donc prévoir l'erreur des résultats

évaluation qualitative ?

# Exercice 3 : Réduction en dimension

1) La relation est :

$$\text{information conservé} = \frac{\overbrace{\lambda_1 + \lambda_2 \ldots}^{\text{autant que la taille de la dimension ciblée}}}{\sum_{i=1}^{m} \lambda_i}$$

1,5

oui, on peut aussi le formuler sous équation

2) La ligne 8 sert à trier les valeurs propres. Cette ligne nous sert à sélectionner les axes d'inerties maximum.
On sélectionne ceux de valeur propre la plus importante.

0,5

3) Les lignes 9 à 10 servent à sélectionner les vecteurs propres qui correspondent aux plus grande valeur propre.

2 Comme d=1, on prends le plus grand $\lambda$, soit $2{,}033e{-}02$, indice 3.
On sélectionne donc la dernière ~~ligne~~ colonne de VP,

0,5

donc $W = \begin{pmatrix} 0{,}0062125 \\ 0{,}9275136 \\ 0{,}3737378 \end{pmatrix}$

4) C'est celui qui correspond à la plus petite valeur propre, soit la première colonne de VP.

* intercalaire

1,5

5) 04 permet de recentrer la VT,
05 permet de réduire, c'est-à-dire de faire en sorte que le nuage de point ne soit plus influencé par les unités de mesure choisies.

0,5

0,5 6) $W = \begin{pmatrix} 0{,}56484 \\ 0{,}59295 \\ 0{,}57390 \end{pmatrix}$

7) * intercalaire

l'anonymat figurant sur la copie principale à reporter : ..............

Rolin groupe 7

## Exo 3

4) suite : Comme la première ligne de ce est 0,0062125, on ne tient quasiment pas compte de la taille. ou

Le vecteur propre contenant le moins d'informations, $\begin{pmatrix} 0,9999628 \\ -0,0080074 \\ 0,0032353 \end{pmatrix}$,
est uniquement orienté ~~par rap~~ dans le "sens de la taille" -> plus grand axe d'inertie pour la taille. ou
Le meilleur est orienté plus vers le poids, et le second meilleur plus vers foot size.

7) Le problème est résolu, on ne dépends plus de l'unité de mesure.
Le meilleur vecteur propre est $\begin{pmatrix} 0,56484 \\ 0,59295 \\ 0,57390 \end{pmatrix}$, à la projection tout est pris en compte avec ce vecteur

1

TB