

Nom :

Prénom :

Groupe :

### 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  $\epsilon$ .

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

$$\text{distance} = |x_{current} - x_{init}|$$

Tant que (distance >  $\epsilon$ )

$$x_{init} = x_{current}$$

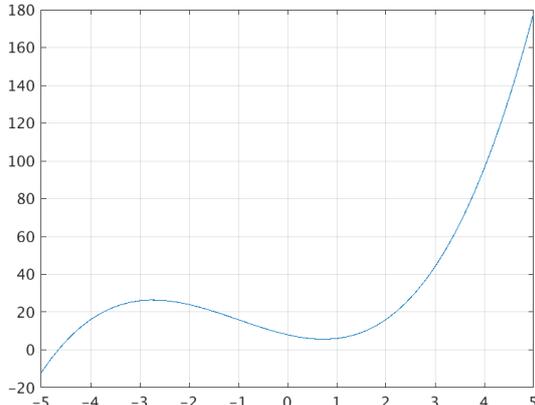
$$x_{current} = x_{init} - nu \times f'(x_{init})$$

$$\text{distance} = |x_{current} - x_{init}|$$

1. Expliquez l'algorithme en répondant aux questions suivantes :

- Où intervient le signe de la dérivée ?
- Quelle variable contient la position du minimum après exécution de l'algorithme ? Quelle est la valeur du minimum ?
- Quelle est l'influence du paramètre  $\epsilon$  ?
- 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 :



- 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.

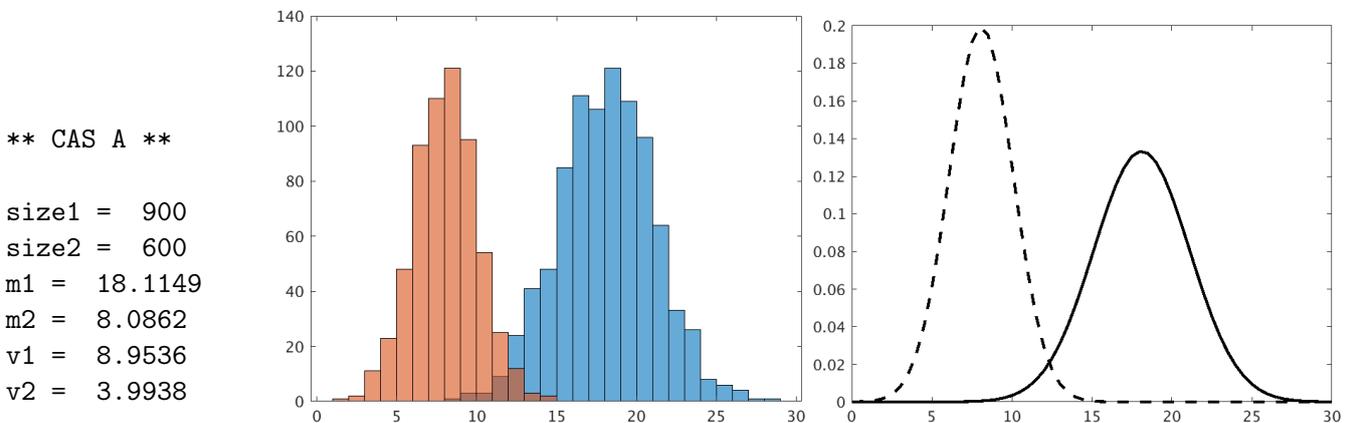
---

```

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)), '—');

```

---



- (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).

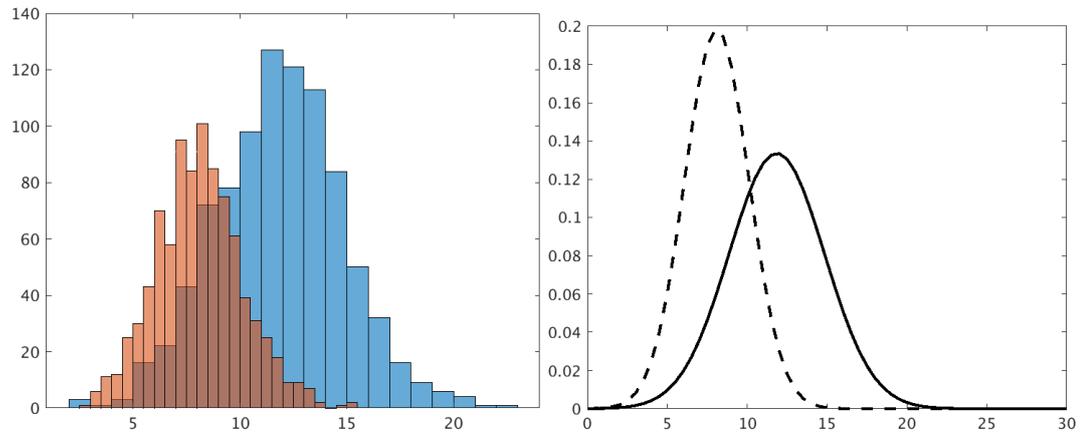
- (c) Le seuil de décision pour le classifieur MAP sera-t-il supérieur ou inférieur à  $x_{seuil}$  ? Expliquez.  
 (d) Au vu des informations fournies, pouvez-vous prévoir la qualité des résultats des classifieurs ML et MAP entraînés avec ces données ? Expliquez.

### 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 :

```
01. H = [ 1.82, 82, 30; 1.80, 86, 28; 1.70, 77, 30; 1.80, 75, 25];
02. F = [ 1.52, 45, 15; 1.65, 68, 20; 1.68, 59, 18; 1.75, 68, 23];
03. VT = [H; F] ;
06. M = cov(VT);
07. [VP, Lambda] = eig (M)
08. [SortedLambda, indices] = sort(diag(Lambda), 'descend');
09. d = 1;
10. indices = indices(1:d);
11. 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.