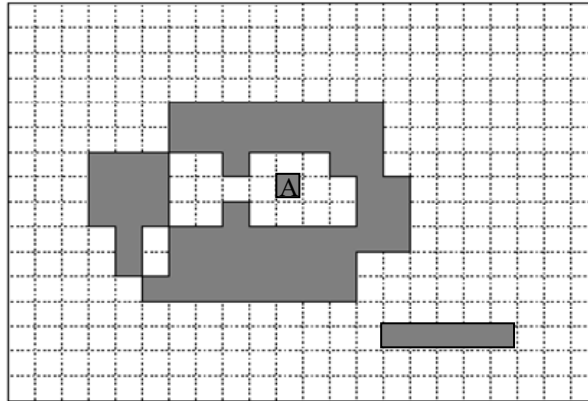


TD4 : Segmentation – Interprétation - Décision

Topologie, Codage de Freeman, Clustering, ...

Exercice n°1 : Topologie

Sur l'image suivante on considère que les niveaux de gris des pixels ont été inversés pour des raisons d'impression (le fond est noir en réalité).



- 1- Tracer la frontière des objets externes en utilisant une méthode de suivi de contours en 4-voisinage et en 8-voisinage.
- 2- Donner le nombre d'objets et la relation d'imbrication de cette image en 4 et 8-voisinage en précisant le nombre de fils des objets externes (qui n'ont pas de parent) en 8-voisinage.
- 3- Donner la chaîne de codes de Freeman dans le cas du 4-voisinage et du 8-voisinage.
- 4- A partir de la chaîne de codes, calculer dans les deux cas le périmètre et la surface des objets en pixels en supposant que l'objet A est sur la ligne i.
- 5- Si on considère que l'objet réel A est un défaut rectangulaire de 2mm de large sur 3mm de haut, donner le périmètre et la surface en mm des objets de l'image.

Exercice n°2: Reconnaissance de formes

On souhaite effectuer la classification d'Iris en fonction de deux paramètres. Les données fournies dans le tableau 1 de l'annexe RDF correspondent à un ensemble de caractéristiques calculées sur 3 types d'Iris après la phase de segmentation.

- 1- On choisit dans un premier temps de faire une représentation des classes en intension en bornant chaque caractéristique indépendamment l'une de l'autre.
 - 1.1. Donner les valeurs des bornes inf. et sup. pour les classes ω_1 et ω_2 .
 - 1.2. Tracer sur le graphique 1 les frontières des classes ainsi obtenues.
 - 1.3. Donner la matrice de confusion. Quel est la justesse (accuracy) obtenue ?
 - 1.4. Comment pourrait-on modifier les frontières de classes pour améliorer le résultat ?
- 2- On choisit maintenant une représentation des classes en extension.
 - 2.1. Donner le vecteur représentatif de chacune des classes.
 - 2.2. Tracer, sur le graphique 1, les formes de classes obtenues ainsi que la frontière inter-classe, dans le cas où on choisit la distance euclidienne pour la classification.
 - 2.3. Quelle forme aurait le vecteur caractéristique si on choisit un modèle Gaussien
- 3- Quelle mesure de distance peut-on choisir pour que la forme des classes soit elliptique ? Qu'est-ce que cela changerait dans la classification ?
- 4- Que peut-on dire sur les échantillons de la classe ω_3 ? Quel problème rencontre-t-on si on applique les méthodes de classification précédentes ?

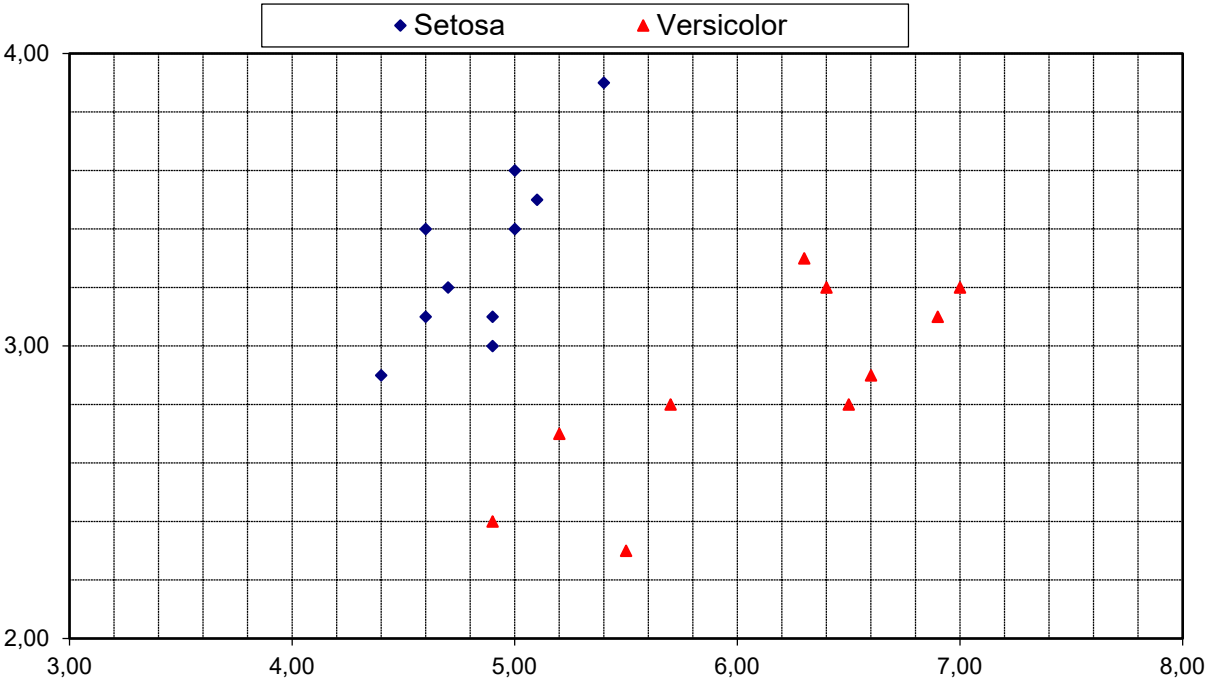
TD4 : Segmentation – Interprétation - Décision
Topologie, Codage de Freeman, Clustering, ...

Annexe RDF :

Tableau 1 :

	Iris-versicolor : ω1		Iris-setosa : ω2		Iris-Virginica : ω3	
	Param 1	Param 2	Param 1	Param 2	Param 1	Param 2
	7,00	3,20	5,10	3,50	6,30	3,30
	6,40	3,20	4,90	3,00	5,80	2,70
	6,90	3,10	4,70	3,20	7,10	3,00
	5,50	2,30	4,60	3,10	6,30	2,90
	6,50	2,80	5,00	3,60	6,50	3,00
	5,70	2,80	5,40	3,90	7,60	3,00
	6,30	3,30	4,60	3,40	4,90	2,50
	4,90	2,40	5,00	3,40	7,30	2,90
	6,60	2,90	4,40	2,90	6,70	2,50
	5,20	2,70	4,90	3,10	7,20	3,60
Moyenne	6,10	2,87	4,86	3,31	6,57	2,94
Ecart-Type	0,73	0,34	0,29	0,31	0,80	0,34

Graphique 1 :



TD4 : Segmentation – Interprétation - Décision
Topologie, Codage de Freeman, Clustering, ...

Exercice n°3 : K-means avec lot de données Iris

- 1- Charger le lot de données UCI iris : ***Iris=load('iris.txt')*** ;
 Construire le vecteur de sortie : ***Y= Iris(: , 1) + 1*** ; (la fonction kmeans numérote les clusters à partir de 1) ;
 Construire le lot de données d'entrée : ***X=Iris(:, 4:5)***; (attributs Longueur et Largeur des pétales).
- 2- Réaliser un clustering en 3 classes avec l'algorithme des K-means (***Yp=kmeans(X,3)*** ;). Visualiser les clusters obtenus.
- 3- Calculer la matrice de confusion issue de cette partition (***CF=confusionmat(Y,Yp)***). Quel problème peut-on rencontrer ?
- 4- Modifier les conditions initiales (***Yp=kmeans(X,3, 'Start', C_i, 'EmptyAction','drop')***) avec :
 $C_1 = [0 \ 0; 0 \ 0; 0 \ 0]$ / $C_2 = [0 \ 0; 2.5 \ 5; 5 \ 10]$ / $C_3 = [0 \ 0; 2.5 \ 1.5; 5 \ 2.5]$
 Calculer la matrice de confusion dans chaque cas. Que conclure ?
- 5- Evaluer les performances du clustering obtenu :
 Calculer la justesse (accuracy : taux de classification / classes / total)
 Calculer la précision et le rappel.
- 6- Effectuer la classification en utilisant 3 variables en entrée :
(X=Iris(:, 3:5) avec C4 = [3 0 0 ; 0 2.5 1.5; 0 5 2.5]) .
 Calculer les indicateurs de performance (notamment accuracy) que conclure ?
- 7- Effectuer la classification en utilisant 4 variables en entrée :
(X=Iris(:, 2:5) avec C5 = [0 3 0 0 ; 0 0 2.5 1.5; 0 0 5 2.5]) .
 Calculer les indicateurs de performance (notamment accuracy) que conclure ?

Exercice n°4 : Arbre de décision avec lot de données Iris

- 1- Utiliser le module « **Classification Learner** » (menu APPS). Créer une nouvelle session en important les données UCI « iris.txt ». Importer l'ensemble des données, choisir la **variable 1** comme réponse. Configurer le module pour travailler **sans validation**.
- 2- Sélectionner les **variables 2 et 3** (Longueur et Largeur des Sépales) et entraîner deux arbres de décision **fine** et **coarse** avec les paramètres par défaut.
- 3- Visualiser la matrice de confusion correspondante et donner les mesures d'évaluation (justesse, rappel, précision, ...) pour les 2 cas.
- 4- Exporter les arbres de décision créés (T_Fine, T_Coarse) et les visualiser avec la commande ***view(T_Fine.ClassificationTree, 'Mode','graph')***. Que conclure ?
- 5- Créer une nouvelle session en choisissant cette fois une **validation croisée 2/3 – 1/3** (3 folds). Evaluer les modèles obtenus comme lors de la question 3.
- 6- Recommencer la manip précédente avec cette fois une **validation croisée 9/10 – 1/10** (10 folds). Evaluer les modèles obtenus. A quel taux peut-on réellement accorder sa confiance ? Que conclure ?
- 7- En choisissant la bonne méthode de validation, refaire un entraînement en prenant en compte cette fois les **variables 4 et 5** uniquement (Longueur et Largeur des Pétales). Evaluer les modèles obtenus.
- 8- Refaire le test précédent en utilisant cette fois les **variables 3, 4 et 5**.
- 9- Que peut-on attendre comme résultats si on utilise les 4 variables d'entrée ?