

Rapport Projet : Algorithme de coalescence

Réalisé par Zakaridia DIAWARA et Dalila HANOUTI

Ce projet a été réalisé en utilisant le logiciel octave. Les codes sources sont dans le dossier **sources**.

1 – Prise en main du programme

- Vecteur moyenne et matrice de variance-covariance

Classe	matrice variance - covariance	Vecteur moyenne
1	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 4 \\ 4 \end{pmatrix}$
2	$\begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$	$\begin{pmatrix} -4 \\ -4 \end{pmatrix}$
3	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$

- L'algorithme de coalescence a été codé dans le fichier **coalescence.m**

- L'exécution de la fonction **prise_en_main()** permet de visualiser le résultat des individus obtenu à partir de l'application de l'algorithme de coalescence. Le choix du centre de gravité se fait de façon aléatoire. Le taux d'erreur par rapport à la classification est également retourné.

- Après exécution de l'algorithme de coalescence et affichage du résultat, nous observons des erreurs de classifications dues à la dispersion des individus au sein des classes (notamment la dispersion des individus dans la partition optimale ayant la classe numéro 2).

2 - Influence du point de l'initialisation

- La visualisation planaire des individus grâce a la fonction **visualiser(«td2_td1.txt»)** nous permet de constater que ceux-ci peuvent être regrouper en deux classes.

- En regroupant les individus par application de l'algorithme de coalescence, nous observons que la répartition obtenu ne nous convient pas a chaque fois.

Le problème se situe lors de l'initialisation des centres de gravité. En effet choisir les deux centres gravités parmi les individus qui se ressemblent le plus, conduit à une partition non satisfaisante.

- Pour choisir des centres de gravité permettant d'obtenir une partition satisfaisante, nous choisissons deux individus en fonction de la moyenne (m) et de l'écart-type (e) de la deuxième mesure. Les deux valeurs devront respecter les contraintes suivantes :

- * La première valeur $e + m$
- * La seconde valeur $m - e$

- Pour visualiser le résultat, il faut exécuter la fonction **influence_pt_init()**. Le choix du bon centre de gravité se fait à travers la fonction **gravite(x,k)** (dans le fichier **gravite.m**)

Voir screenshot dans **images/gravite_corrige.png** (partition obtenu chaque fois après proposition de centre de gravité) et **images/gravite_probleme.png** (partition obtenu quelques fois avec choix aléatoire de centre de gravité)

3 - Influence de la métrique

La visualisation planaire des individus grâce a la fonction **visualiser**(«td2_td2.txt») .

On sait que l'écart type de la mesure 1 est de 0.5 et que l'écart type de la mesure 2 est de 2 . En appliquant la solution précédente nous obtenant un résultat insatisfaisant , donc la métrique a bien influencer sur le résultat.

La solution proposé est consiste à calculer la métrique de la façon suivante:

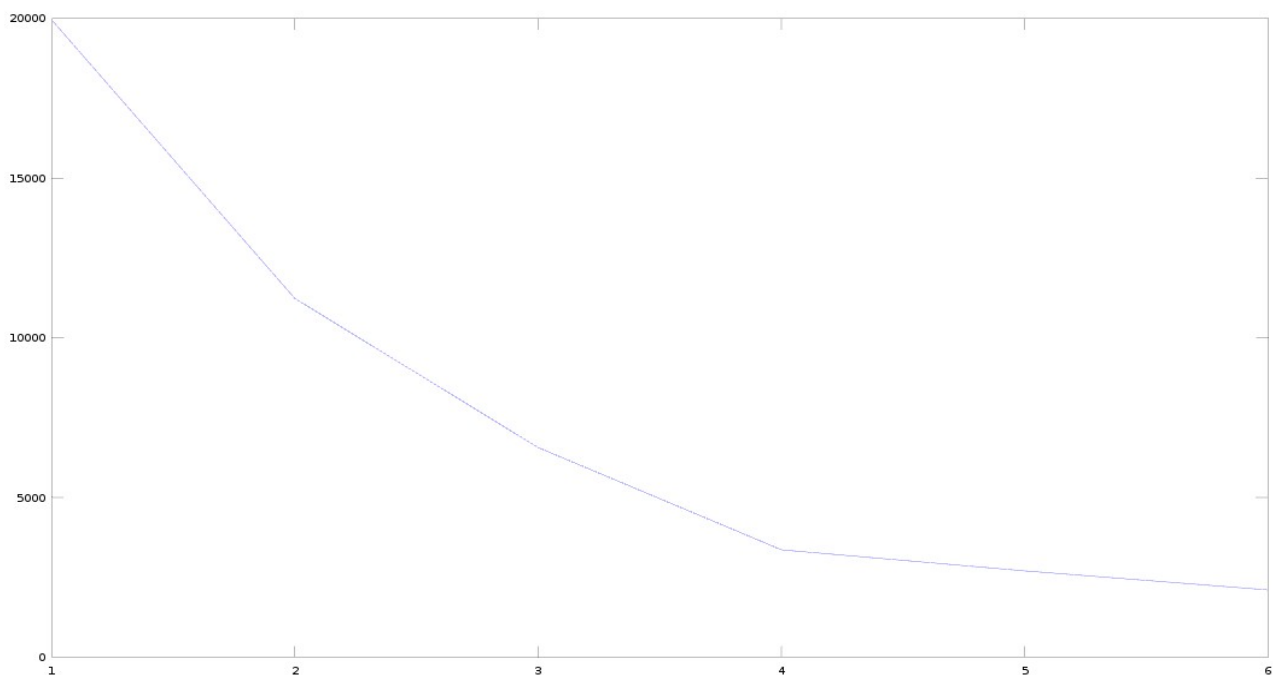
$$M(1,1)=1/\text{VAR}(x(1,:)).$$

$$M(2,2)=1/\text{VAR}(x(2,:)).$$

4 - Choix du nombre de classes

L'algorithme de coalescence vise à minimiser la valeur de l'inertie intra-classe.

D'après la courbe de représentation de cette valeur en fonction du nombre de classe, nous constatons que celle-ci décroît très peu à partir du nombre de classe $K = 4$. Nous pouvons donc conclure que le nombre de classe de l'ensemble d'apprentissage est 4.



Courbe représentant nombre de classe en fonction de l'inertie.