Rapport TP Coalescence

Zakaridia - Dalila

12 octobre 2015

Ce projet a été réalisé en utilisant le logiciel octave.

1 Prise en main des programmes

— Tableau récapitulant le vecteur moyenne et matrice de variance-covariance pour les trois classes.

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

— La fonction $[x\ m\ clas] = initialiser()$ permet de générer aléatoirement les données de test et d'initialiser la partition optimale. Elle renvoie également la métrique associée aux données de test. En utilisant la fonction $affiche\_classe(x, clas)$, nous obtenons la représentation graphique suivante représentant celle de la partition optimale.

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

— Pour tester l'algorithme, nous procédons comme suite :

  — Initialisation des données : $[x\ m\ clas] = initialiser()$

  — Exécution de la fonction $[clas2\ g] = lancerAleatoire(x, k, m)$: Dans cette fonction, nous choisissons aléatoirement les centres de gravité initiaux puis on exécute l'algorithme de coalescence. $clas2$ contient le résultat de la classification et $g2$ le centre de gravité par classe.

— 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

- Chargement des données : $[x\ m] = charger("td2\_d1.txt")$

- Visualisation planaire des données : $visualiser(x)$

- La visualisation planaire des individus nous permet de constater que ceux-ci peuvent être regrouper en deux classes.
  En regroupant donc ces 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'ecart-type (e) de la deuxième mesure. Les deux valeurs devront respecter les contraintes suivantes :
  - La première valeur $\geq e + m$
  - La seconde valeur $\leq m - e$

3 Influence de la métrique

- Chargement des données : $[x\ m] = charger("td2\_d.txt")$

- Visualisation planaire des données : $visualiser(x)$

- La visualisation planaire des individus nous permet de constater que ceux-ci peuvent être regrouper en deux classes.
  En regroupant donc ces 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.Et pour palier a ce probleme nous proposons une solution qui consiste a utilier la meyenne de la premiere et la deuxieme mesure pour calculer les deux mesures associers aux deux centre de gravite G1 et G2 .
  - Calculer la moyene de la premiere mesure Moym1 .
  - Calculer la moyene de la deuxieme mesure Moym2.
    - G1 (Moym1 , (min(x(2, :)) + moyM2)/2)).
    - G2 (Moym2 , (max(x(2, :)) + moyM2)/2).

  le resultat obtenu est très satisfaisant.
- avec l'ecart type de la mesure 1 est de 0.1 et que l'ecart type de la

mesure 2 et de 2 on appliquant la solution precedente nous obtenant un resultat insatisfaisant ,donc la metrique a bien influencer sur le resulat. La solution proposé est consiste à calculer la metriqe de la façon suivante :
— M(1,1)=1/VAR(x(1, :)).
— M(2,2)=1/VAR(x(2, :)). Le resultat obtenu est tres satisfaisant .

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.