Locality Sensitive Discriminant Analysis

Quand on considère un ensemble de données, on peut faire la distinction entre la structure globale et une structure locale du nuage de points. Dans le cadre de l’analyse discriminante, la LDA échoue à découvrir cette structure locale, qui est généralement plus importante que la structure globale lorsqu’il y a peu de données d’entrainement.

On cherche à pallier ce défaut par une méthode d’analyse discriminante sensible à la localité (LSDA), présentée par cinq chercheurs en 2007 lors de l’International Joint Conference on Artificial Intelligence.

(inclure la ref dans le diapo)

Ref : D. Cai, X. He, K. Zhou, J. Han, and H. Bao, “Locality sensitive discriminant analysis,” in International Joint Conference on Artificial Intelligence (IJCAI’07), 2007.

Là où la LDA va chercher à représenter les données en fonction des axes les plus discriminants en regardant les statistiques globales (moyennes et variances), cette autre méthode va chercher à maximiser la variance entre les points de classes différentes à chaque voisinage local, et pour ce faire on introduira une notion de voisinage.

On a souvent une configuration où les données présentent entre elles de nombreuses relations de corrélation, càd qu’elles vivent dans un sous-espace de dimension inférieure à l’espace ambiant. Pour se ramener à un cas où les variables sont décorrélées, on peut par exemple procéder à une ACP.

On se place donc dans le cas où les données vivent dans ou proche de l’espace ambiant (pas ou peu de relations de corrélation). On cherche une matrice de transformation linéaire A telle que Y = XA où Y est la matrice de données projetées sur un espace de dimension d et X est la matrice de données initiale.

Pour modéliser les structures locale et globale, on construit deux graphes des plus proches voisins, l’un concernant les points avec le même label et l’autre pour les points avec un label différent.