Master 1

jeudi 6 mai 2021

Contrôle
Visualisation de données
(sans documents)

Exercice 1 : *question de cours*

Compléter le tableau ci-dessous pour caractériser chacune des méthodes vues en cours, en précisant le critère utilisé, et la nature de la méthode d'optimisation de ce critère.

	Critère	**Méthode d'optimisation**
ACP		
LDA		
MDS		
Laplacian Eigenmaps		
t-SNE		

Exercice 2 : *Locally Linear Embedding*

On rappelle que le principe de la méthode Locally Linear Embedding (LLE) est d'approcher chaque point $X_i \in \mathbb{R}^n$ par une combinaison linéaire de ses K plus proches voisins X_{ij}. Si l'on note $\tilde{X}_i$ l'approximation de X_i et $w_i = (w_{i1} \quad \cdots \quad w_{ik})^T$, alors on a la relation suivante

$$\tilde{X}_i = \sum_{j=1}^{k} w_{ij} X_{ij} \qquad avec \quad \sum_{j=1}^{k} w_{ij} = 1_k^T w_i = 1 \qquad (1)$$

et l 'erreur d'approximation totale des m points X_i s'écrit

$$\min_{w_1, w_i \dots w_m} \sum_i \left\| X_i - \sum_{j=1}^{k} w_{ij} X_{ij} \right\|^2 \qquad (2)$$

On cherche alors les points projetés $Y_i \in \mathbb{R}^d$ avec $d < n$, qui satisfont la même contrainte d'approximation par les K plus proches voisins Y_{ij} dans l'espace projeté, c'est à dire les Y_i qui minimisent l'erreur suivante

$$\min_{Y_1, Y_i, \dots, Y_m} \sum_i \left\| Y_i - \sum_{j=1}^{k} w_{ij} Y_{ij} \right\|^2 \qquad (3)$$

Question 1 :

Si l'on note $N_i = (X_{i1} \quad \cdots \quad X_{ik})$ la matrice de dimension (n, k) contenant les K plus proches voisins de X_i, ***montrez que la relation suivante est vérifiée, et donner l'expression de G_i***

$$\varepsilon_i = \left\| X_i - \sum_{j=1}^{k} w_{ij} X_{ij} \right\|^2 = {w_i}^T G_i w_i \qquad (4)$$

Question 2 :

On détermine les vecteurs de poids w_i qui minimisent l'erreur d'approximation ε_i sous la contrainte $1_k^T w_i = 1$. Si l'on note λ_i le multiplicateur de Lagrange associé à cette contrainte, alors on doit minimiser le Lagrangien suivant

$$\mathcal{L}_i = {w_i}^T G_i w_i + \lambda_i (1 - 1_k^T w_i) \qquad (5)$$

En dérivant par rapport à w_i, ***montrer que***

$$w_i = \frac{1}{2} {G_i}^{-1} \lambda_i 1_k \qquad (6)$$

En introduisant la contrainte $1_k^T w_i = 1$, **déduire que**

$$\lambda_i = \frac{2}{1_k^T {G_i}^{-1} 1_k} \qquad (7)$$

Déduire de 6) et 7) l'expression finale des vecteurs w_i ci-dessous

$$w_i = \frac{{G_i}^{-1} 1_k}{1_k^T {G_i}^{-1} 1_k} \qquad (8)$$

Question 3 :

On cherche maintenant à déterminer les points projetés Y_i en minimisant l'expression (3). On montre que cette expression peut s'écrire

$$\sum_i \|Y^T 1_i - Y^T \widetilde{w}_i\|^2 = \left\|Y^T\left(I - \widetilde{W}^T\right)\right\|_F^2 \qquad (9)$$

Donner l'expression, et préciser la dimension des matrices et vecteurs ci-dessous

$$Y^T \qquad 1_i \qquad \widetilde{w}_i \qquad \widetilde{W}^T$$

Question 4 :

On note M la matrice ci-dessous

$$\mathrm{M} = \left(I - \widetilde{W}^T\right)^T \left(I - \widetilde{W}^T\right) \qquad (10)$$

Préciser la dimension de cette matrice

Les points Y_i rassemblés dans la matrice Y , qui minimisent l'expression (9) sous les contraintes $\frac{1}{m} Y^T Y = I$, $Y^T 1_m = 0_d$, sont déduits des vecteurs propres de la matrice M.

Préciser la dimension de chaque vecteur propre de M.

Combien y en a-t-il au maximum ?

En excluant celui de valeur propre nulle, quels sont les vecteurs propres qu'il faut retenir pour minimiser le critère lorsque que l'espace projeté recherché est de dimension d ?

Question 5 :

On se place maintenant dans la situation où l'on cherche à exploiter les résultats de la méthode LLE optimisée sur un premier ensemble de points, pour déduire les projections de nouveaux points sans procéder à un réapprentissage.

Chaque nouveau point $X^t \in \mathbb{R}^n$ n'appartenant pas à l'ensemble d'apprentissage est projeté en $Y^t \in \mathbb{R}^d$ en exploitant uniquement les points $X_i, i = 1, \dots, m$ de l'ensemble d'apprentissage et leurs projections Y_i déterminées lors de l'apprentissage.

Proposer une méthode pour déterminer $\boldsymbol{Y^t} \in \mathbb{R}^{\boldsymbol{d}}$, et en préciser les étapes de calcul.