TD 3

On admet le résultat suivant, l'une des conséquences les plus importantes de la SVD :

Théorème d'Eckart-Young

Soit $\mathbf{A} \in \mathbb{R}^{n\times d}$. La norme de Frobenius de $\mathbf{A}$ est définie par $\|\mathbf{A}\|_F^2 = \sum_{i,j} A_{ij}^2$. Soit E_r l'ensemble des matrices $\mathbf{B} \in \mathbb{R}^{n\times d}$ telles que $\text{rang}(\mathbf{B}) \leq r \leq \min(d, n)$. On écrit la décomposition SVD de $\mathbf{A} = \sum_{i=1}^{\min(n,d)} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$, où $\sigma_1 \geq \sigma_2 \geq ....$ Alors la meilleure approximation de rang maximal r de $\mathbf{A}$ définie par

$$\min_{\mathbf{B}\in E_r} \|\mathbf{A} - \mathbf{B}\|_F^2$$

est donnée par :

$$\mathbf{B}^\star = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

Exercice 1 (Lemme : Décomposition matricielle)
Soit $\mathbf{A} \in \mathbb{S}_d$ une matrice symétrique semi-définie positive de taille d. On suppose que son rang est égal à r. Montrez qu'il existe $\mathbf{Z} \in \mathbb{R}^{d\times r}$ telle que :

$$\mathbf{A} = \mathbf{Z}\mathbf{Z}^\top$$

.

Exercice 2 (Classical MDS)
Soit $\mathrm{X} \in \mathbb{R}^{n\times d}$ une matrice des données en dimension d d'un vecteur aléatoire centré. Classical MDS cherche la meilleure représentation en dimension $q < d$ ($\mathrm{Z} \in \mathbb{R}^{n\times q}$) qui préservent les produits scalaires $\langle \mathbf{x}_i, \mathbf{x}_j\rangle$. Ainsi on cherche à résoudre :

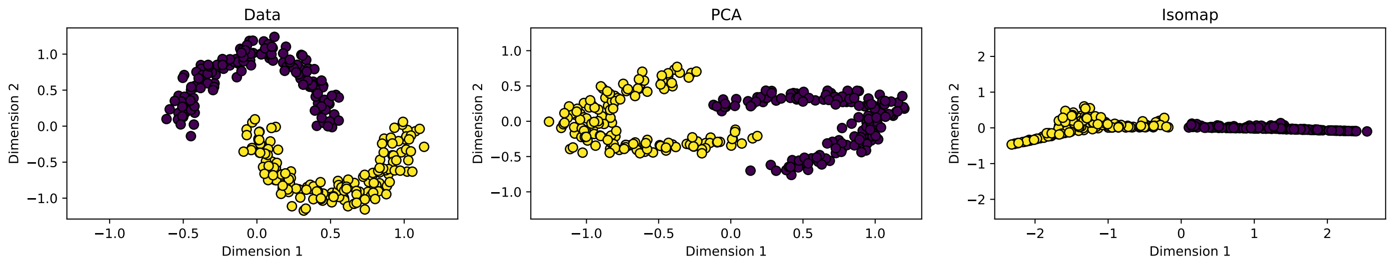
$$\min_{\mathbf{z}_1,\dots,\mathbf{z}_n\in\mathbb{R}^q} \sum_{i,j}^{n} (\mathbf{x}_i^\top \mathbf{x}_j - \mathbf{z}_i^\top \mathbf{z}_j)^2$$

1. Comment peut-on écrire le problème d'optimisation ci-dessus en fonction des matrices X et Z et de la norme de Frobenius ?
2. Montrez que le problème Classical MDS est équivalent à une approximation de faible rang et en déduire que la solution est donnée par les q premiers vecteurs propres de la matrice $\mathrm{X}\mathrm{X}^\top$ multipliés par des scalaires à déterminer.
3. En effectuant une SVD de X, déterminez les valeurs propres et les vecteurs propres de $XX^\top$ et $X^\top X$ en fonction des éléments de la SVD de X.
4. En déduire que les projetés $\mathbf{z}_i$ sont les mêmes que les projetés $\text{PCA}_q(\mathbf{x}_i)$.

Exercice 3 (Classical MDS : double centering)
Soit $\mathrm{X} \in \mathbb{R}^{n\times d}$ une matrice des données en dimension d d'un vecteur aléatoire centré, on suppose donc que la moyenne empirique est nulle. On suppose que l'on a pas accès à la matrice X ni aux produits scalaires $\langle \mathbf{x}_i, \mathbf{x}_j\rangle$. Mais on a la matrice des distances Euclidiennes $D_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2$. On définit $A \stackrel{\text{def}}{=} XX^\top$.

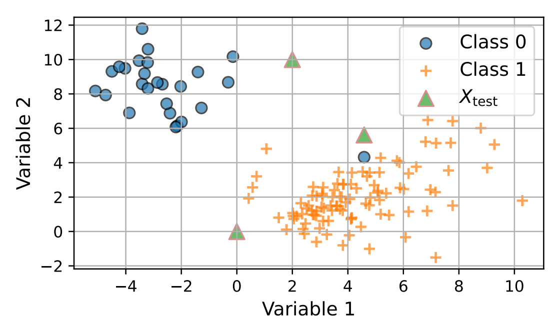
1. Écrivez d_{ij} en fonction de A_{ij}, A_{ii} et A_{jj}.
2. Calculez $\sum_{i=1}^n d_{ij}$ et $\sum_{j=1}^n d_{ij}$ et $\sum_{i,j}^n d_{ij}$ en fonction de A_{jj}, A_{ii} et trace(A).
3. En déduire A_{ij} en fonction des termes de D.
4. Montrez que $A = -\frac{1}{2}HDH$ où $H = \mathrm{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^\top$.
5. On appelle H une matrice de centrage. Pourquoi ?
6. En déduire un algorithme Classical MDS à appliquer sur une matrice des distances $D \in \mathbb{R}^{n\times n}$.

Exercice 4 (Analyse de données)
On observe 200 observations en deux dimensions. On applique la PCA et Isomap avec 10 voisins en deux dimensions dans le but d'essayer d'isoler les deux clusters à part et de pouvoir les séparer avec une droite.



1. La PCA semble avoir échangé les deux variables. Expliquez pourquoi.
2. La PCA n'a pas réussi à séparer les deux clusters. Qu'en est-il de Classical MDS ?
3. Et si on utilisait MDS avec les distances Euclidiennes ?
4. Expliquez la capacité d'Isomap a rassembler les deux classes.

Exercice 5 (Prédiction simple)
On observe n observations en deux dimensions données par la matrice $\mathrm{X} \in \mathbb{R}^{n\times 2}$ pour lesquels on a des labels binaires $y_1, \dots, y_n$. On observe également 3 échantillons "X_{test}" pour lesquels on n'a pas accès aux labels. Voici la visualisation entière de ces données.



1. Si on tranforme ces données avec une PCA avec deux composantes, la représentation des données serait-elle différente des données d'origine ? Justifiez votre réponse.
2. On souhaite utiliser les données labélisées pour prédire les labels des 3 observations de test. En se basant sur la figure ci-dessus, déterminez les labels des trois échantillons de gauche à droite en utilisant le modèle k-Nearest-Neighbors avec $k = 1$ et $k = 2$ et $k = n$.
3. À quoi correspond la fonction de prédiction k-NN $\varphi : \mathbf{x} \in \mathbb{R}^2 \mapsto \mathbf{y} \in \{0, 1\}$ si $k = n$?