
PROJET 1

JEAN-THOMAS BAILLARGEON
CHRISTOPHER BLIER-WONG
POUR LE COURS STT-7330
MÉTHODE D'ANALYSE DES DONNÉES

PRÉSENTÉ LE 25 MARS 2018 À LA PROFESSEURE

ANNE-SOPHIE CHAREST

*Département de mathématiques et de statistiques
Faculté des sciences et de génie
Université Laval*



UNIVERSITÉ
LAVAL

JEAN-THOMAS BAILLARGEON
CHRISTOPHER BLIER-WONG
FACULTÉ DES SCIENCES ET DE GÉNIE
ÉCOLE D'ACTUARIAT
UNIVERSITÉ LAVAL
HIVER 2018

Table des matières

1	Introduction	2
2	Fonction noyau	2
3	PCA avec noyau	3
3.1	Normalisation des données	3
4	Application pratique	4
5	Conclusion	6

1 Introduction

L'objectif de l'analyse par composantes principales est d'obtenir une représentation des données dans un espace plus restreint en conservant la plus grande quantité d'information possible. Plus précisément, on considère les combinaisons linéaires des variables mesurées pour l'espace restreint. Par contre, on observe souvent que la relation entre les données n'est pas linéaire. L'analyse par composantes par noyau généralise ce modèle pour des combinaisons non-linéaires des attributs. Au lieu de faire une décomposition par valeurs et vecteurs propres sur la matrice de covariance des données centrées

$$\Sigma = \text{Var}(X) = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T,$$

on fait une décomposition par valeurs et vecteurs propres sur la matrice de covariance des données projetées sur un nouvel espace d'attributs.

2 Fonction noyau

L'idée de l'espace des attributs est de projeter les données originales par une fonction non linéaire vers un nouvel espace. Formellement, on a

$$\begin{aligned} \Phi : \mathbb{R}^N &\rightarrow \mathcal{F} \\ \mathbf{x} &\mapsto \Phi(\mathbf{x}) \end{aligned}$$

où les données $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^N$ est projeté vers un espace d'attributs \mathcal{F} [Muller et al., 2001]. Souvent, la dimension de \mathcal{F} est beaucoup plus élevée que l'espace originale. L'apprentissage statistique peut maintenant être fait sur les données $(\Phi(\mathbf{x}_1), y_1), (\Phi(\mathbf{x}_2), y_2), \dots, (\Phi(\mathbf{x}_n), y_n)$.

Le produit scalaire entre deux espaces d'attributs peut être reformulé en terme d'une fonction noyau k par

$$k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})).$$

Dans plusieurs problèmes d'apprentissage, le "truc du noyau" permet d'éviter de calculer directement les nouvelles données $\Phi(\mathbf{x}_n)$. En effet, pour certains algorithmes d'apprentissage, il n'est pas nécessaire d'avoir toutes les données car on peut reformuler les équations de mise à jour par le produit scalaire entre différentes données et ainsi les remplacer par la fonction de noyau. Des exemples de noyaux communément utilisés sont présentés dans la table 1.

Nom	$k(\mathbf{x}, \mathbf{y})$
Gaussien (RBF)	$\exp\left(\frac{-\ \mathbf{x}-\mathbf{y}\ ^2}{c}\right)$
Polynomial	$((\mathbf{x} \cdot \mathbf{y} + \theta))^d$
Sigmodoidal	$\tanh(\kappa(\mathbf{x} \cdot \mathbf{y}) + \theta)$
Multiquadrique inversé	$\frac{1}{\sqrt{\ \mathbf{x}-\mathbf{y}\ ^2 + c^2}}$

TABLE 1 – Noyaux communs

3 PCA avec noyau

Dans le contexte de l'analyse par composantes principales linéaire, on trouvait les valeurs et vecteurs propres qui correspondaient à la matrice de covariance

$$\text{Var}(X) = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T.$$

[Schölkopf et al., 1997] proposent de répéter cette analyse sur une transformation des données originales, i.e. trouver les valeurs et vecteurs propres de

$$\text{Cov}(\Phi(X)) = \frac{1}{N} \sum_{i=1}^N \Phi(\mathbf{x}_i) \Phi(\mathbf{x}_i)^T. \quad (3.1)$$

Ensuite, en appliquant le "truc du noyau" présenté dans la section (2), on évite de calculer explicitement les données $\Phi(\mathbf{x})$, il suffit de calculer la matrice des noyaux.

3.1 Normalisation des données

Dans le contexte de l'analyse par composantes principales, on conseil souvent de centrer et réduire les données. Par contre, dans le contexte de projection des données dans l'espace \mathcal{F} et profiter du truc du noyau, on ne peut pas calculer $\tilde{\Phi}(\mathbf{x}_i) = \Phi(\mathbf{x}_i) - \frac{1}{N} \sum_{i=1}^N \Phi(\mathbf{x}_i)$. Soit la matrice K , où

$$K_{ij} = (k(\mathbf{x}_i, \mathbf{x}_j))_{ij}.$$

Il est possible que $\frac{1}{N} \sum_{i=1}^N \Phi(\mathbf{x}) \neq 0$. On doit alors centrer les données selon la passe du p'tit cochon qui tousse suivante :

$$\begin{aligned} \tilde{k}(\mathbf{x}_i, \mathbf{x}_j) &= \tilde{\Phi}(\mathbf{x}_i)^T \tilde{\Phi}(\mathbf{x}_j) \\ &= \left(\Phi(\mathbf{x}_i) - \frac{1}{N} \sum_{l=1}^N \Phi(\mathbf{x}_l) \right)^T \left(\Phi(\mathbf{x}_j) - \frac{1}{N} \sum_{l=1}^N \Phi(\mathbf{x}_l) \right) \\ &= k(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{N} \sum_{l=1}^N k(\mathbf{x}_i, \mathbf{x}_l) - \frac{1}{N} \sum_{l=1}^N k(\mathbf{x}_j, \mathbf{x}_l) + \frac{1}{N^2} \sum_{l,k} k(\mathbf{x}_l, \mathbf{x}_k). \end{aligned}$$

Alors, on peut centrer la matrice \tilde{K} selon

$$\tilde{K}_{ij} = K - \mathbb{1}_N K - K \mathbb{1}_N + \mathbb{1}_N K \mathbb{1}_N,$$

où $(\mathbb{1}_N)_{ij} := \frac{1}{N}$ et éviter de calculer les données $\Phi(\mathbf{x})$.

4 Application pratique

On applique l'analyse par composantes principales et l'analyse par composantes principales avec noyau sur le jeu de données MNIST. Une donnée représente l'intensité de gris entre 0 et 255 des 728 pixels d'une image 28×28 de chiffres entre 0 et 9. Elles ont été récoltées par [LeCun et al., 1998]. Voici un exemple des données :

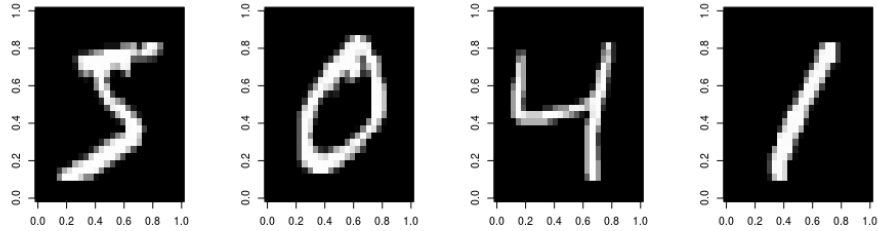


FIGURE 1 – 4 premiers exemples de MNIST

On crée aussi un jeu de données modifié de MNIST, où on applique un bruit gaussien. Voici les mêmes exemples que dans la figure 1 :

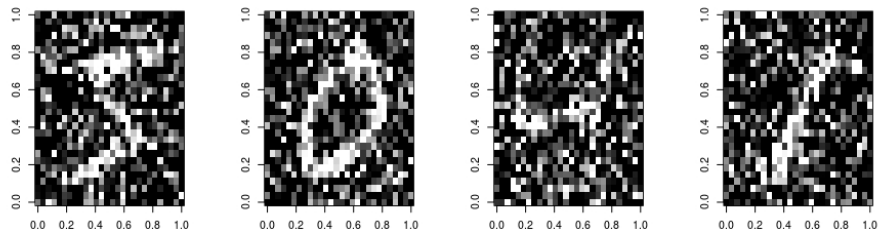
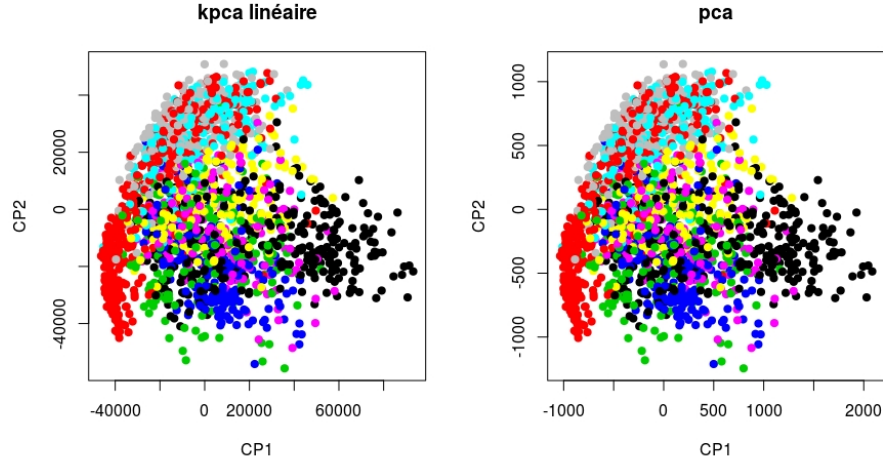


FIGURE 2 – 4 premiers exemples de MNIST avec bruit gaussien

Une analyse par composantes principales est identique à une analyse par composantes principales avec un noyau linéaire. On projette les données selon les deux premières composantes principales. On obtient



Il n'y a pas de flexibilité à cette méthode. On remplace la matrice variance-covariance par différents noyaux et on présente les projections dans la prochaine figure.

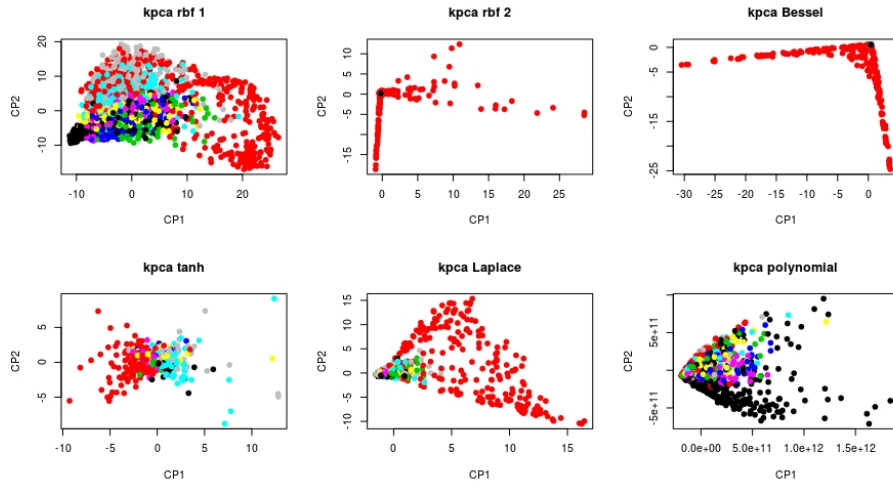


FIGURE 3 – Projection des deux premières composantes principales selon différents noyaux.

On remarque que le choix du noyau a beaucoup d'importance sur la projection. Les données en rouge représentent le chiffre 1 et la plupart des noyaux peuvent séparer les données. Le noyau polynomial est performant pour segmenter les chiffres 0.

On applique ensuite kACP sur les données MNIST bruitées. On obtient

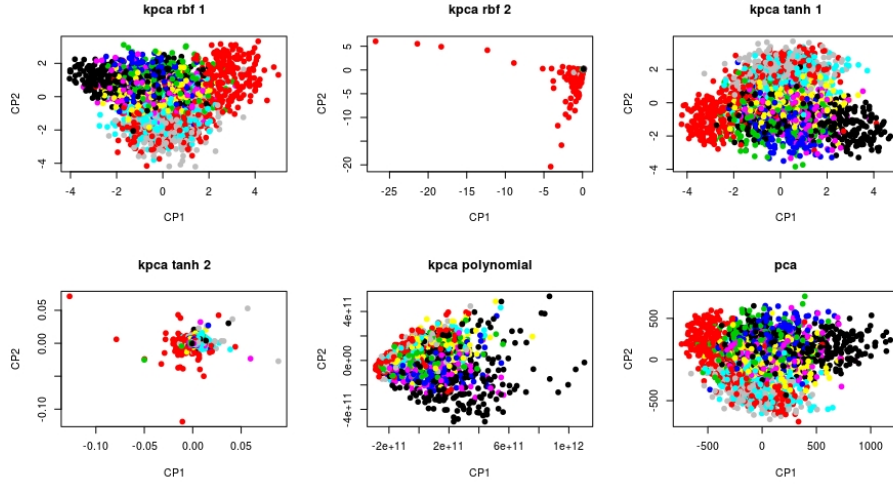


FIGURE 4 – Projection des deux premières CPs selon différents noyaux.

On remarque que certains noyaux sont moins performants pour capturer l'information que sans le bruit, mais que certains sont capables de trouver les relations non-linéaires dans les données bruitées pour retrouver l'information originale.

5 Conclusion

Références

- [LeCun et al., 1998] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324.
- [Muller et al., 2001] Muller, K.-R., Mika, S., Ratsch, G., Tsuda, K., and Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE transactions on neural networks*, 12(2) :181–201.
- [Schölkopf et al., 1997] Schölkopf, B., Smola, A., and Müller, K.-R. (1997). Kernel principal component analysis. In *International Conference on Artificial Neural Networks*, pages 583–588. Springer.