KMEANS and Kernel KMEANS

Ezukwoke K.I[1], Samaneh M.J[2]

[1, 2] Department of Computer Science
Machine Learning and Data Mining
{ifeanyi.ezukwoke, samaneh}@etu.univ-st-etienne.fr
University Jean Monnet, Saint-Etienne, France

Abstract

Kernel principal component analysis (kernel-PCA) is a prominent non-linear extension of one of the most used classiical dimensionality reduction algorithms (PCA-Principal Component Analysis). In this paper, we present an inhaustive comparison between the classical PCA and the kernel-PCA. We use the moon dataset

Keywords
Dimensionality reduction, PCA, kernel-PCA

1 INTRODUCTION

K-Means clustering is a fast, robust, and simple algorithm that gives reliable results when data sets are distinct or well separated from each other in a linear fashion. It is best used when the number of cluster centers, is specified due to a well-defined list of types shown in the data. However, it is important to keep in mind that K-Means clustering may not perform well if it contains heavily overlapping data, if the Euclidean distance does not measure the underlying factors well, or if the data is noisy or full of outliers

References

[1] Pearson K. On lineas and Planes of closest fit to systems of points in space. *Philos Mag A*, 6:559–572, 1901.

[2] Zhu J, Hastie T. Kernel logistic regression and import vector machine. *J Comput Graphic Stat*, 14:185–205, 2005.

[3] Saunders C., Gammerman A., Vovk V., Ridge Regression Learning Algorithm in Dual Variables. *ICML 15th Internationl Conforence on Machine Learning.*, 1998.