Kernel Choice for Unsupervised Learning

Ph. D. Student
Alam M. Ashad
The Graduate University for Advanced Studies

1 Motivation

Kernel methods are increasingly used to solve various problems in statistical science and also in many other research fields. Its offer versatile tools to process, analyze, and compare many types of data and offers state-of-the-art performance in many cases. In machine learning, the field of research devoted to the formal study of learning systems, the positive definite kernels have become wider applicable methods, Hofmann et al.(2008).

During the last one decade the unsupervised learning has become one of the utmost application area of kernel methods. While the choice of kernel in kernel methods is essential, the kernel choice in unsupervised learning has not yet been established. In this paper, an attempt has been made to choice the kernel in kernel canonical correlation analysis and kernel principal component analysis through cross-validation, a popular method of model-selection. We also have studied the numbers of principal component to keep in kernel principal components analysis. We investigate the performance of radial basis kernel(RBK) by simulated data with different models.

2 Unsupervised Learning

We may distinguish between two different types of machine learning likely, supervised learning and unsupervised learning. In unsupervised learning the machine entertain only inputs but obtains neither supervised target outputs nor rewards from its environment.

2.1 Kernel Principal Component Analysis(Kernel PCA)

Kernel PCA is a powerful tool for extracting structure form a high dimensional feature space, nonlinear related to the input variables, proposed by Schölkof, and Somla, (1998). It is computationally similar to classical PCA on n observations with $n \times n$ kernel matrix. Like classical PCA, the number of component to keep for kernel PCA is essential but not yet establish. On other hand the choice of kernel in kernel PCA also not suggested yet.

In kernel PCA, we have different feature space for each kernel and not able to apply cross-validation for kernel choice directly. We would apply cross-validation

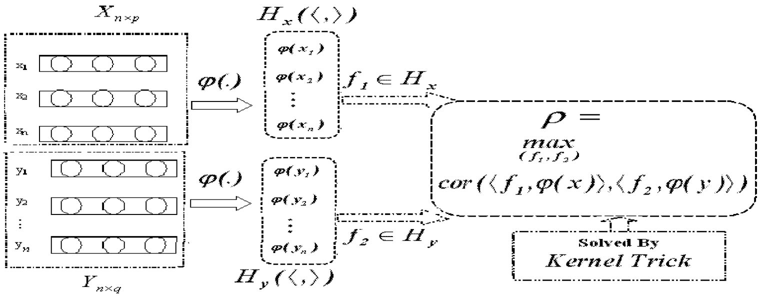

Figure 1: The system of kernel CCA.

to keep number of components and choice of kernel by exact pre-image or approximate pre-image of corresponding input pattern

2.2 Kernel Canonical Correlation Analysis(Kernel CCA)

Kernel canonical correlation is a such type of powerful tool to seek such a method that gives us accurate measurement in case of nonlinear data. This nonlinear technique proposed by Akaho (2000). While kernel CCA has gained popularity as well as implementation in the nonlinear analysis last few years, Bach, et al.(2002); Fukumizu et al. (2007) and Hardoon et al. (2004) but the problem of kernel choice in this method is not fixed yet.

For leave one out cross-validation, we are not able to extract the kernel canonical correlation. Canonical correlation analysis is the generalization of regression analysis with multiple response variables. In this work , we use mean square error like as regression analysis to perform leave one out cross-validation for kernel choice in kernel CCA.

References

Akaho, S. (2000). A kernel method for canonical correlation analysis, *In Proceeding of the 2000 Workshop on Information -Based Induction Science (IBIS2000)* 123-128, Japan.

Bach, F. R., and Jordan,M. I. (2002). Kernel independent component analysis, *Journal of Machine Learning Research* , **3**,1-48.

Fukumizu,K., F.R. Back,F. R., and Gretton, A.(2007). Statistical Consistency of Kernel Canonical Correlation Analysis, *Journal of Machine Learning Research*, **8**, 361-383.

Hardoon, D. R., Szedmak, S., and Shawe-Taylor J. (2004). Canonical correlation analysis: an overview with application to learning methods, *Neural Computation*, **16**, 2639-2664.

Hofmann, T. Schölkof, B., and Somla,J. A.(2008). Kernel Methods in Machine Learning, *The Annals of Statistics*,**36**, 1171-1220.

Schölkof, B., and Somla,J. A.(1998).Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput*, **10**,1299-1319.