THEORETICAL REPORT
IFS-TR-005

USE OF PRINCIPAL COMPONENTS ANALYSIS WITHIN AMBER FOR DIMENSION REDUCTION

ABSTRACT

Neural networks and other supervised learning techniques often encounter difficulties when vectors being input to them are significantly correlated. At best, an input vector with correlated inputs will lead to an overly complex model. In the case of neural networks, this will cause training the network to be slower and also may lead to poor generalisation performance.

In this report, Principal Components Analysis (PCA) is examined as a method of de-correlating input components. As well as de-correlating the inputs, PCA also provides a framework whereby the components are ordered by information content. This allows the technique of dimension reduction of the input vectors with a guarantee that the reduced dimension of the vectors will retain the maximum possible information content (this is information in the Shannon sense). For details of using PCA within Amber consult the *Amber Reference Guide.*

REPORT

A number of authors have suggested Principal Components Analysis (PCA) as a method of reducing the dimension of the input vectors for neural networks in financial and other pattern recognition applications.

When extracting features from a data set it is likely that a good deal of redundant information will be included in the features. This will be particularly true in the case of time series analysis in which several highly correlated time series are being used to attempt to predict future values of another time series. An example of this might be, for instance, the use of the spot USD-DEM rate, 3-month German interest rates and the DAX index to estimate future price movements in a German equity. There will be significant correlation in these time series and as such, any input information that combines lagged values of all of these series will contain much extraneous or redundant information.

This redundant information makes the process of training neural networks more difficult. In particular, it will lead to slower training times and possibly over-complex final architectures. It would be useful to be able to reduce the dimension of the input information through a transform that will maintain the significant information whilst eliminating much of the redundant information and at the same time de-correlating the inputs.

PCA works by finding a subspace of the original input space that preserves the maximum information content (variance) in the original data when it is projected from the original space onto the subspace. The projection of the original data onto the subspace is then used as the input to the neural net. It is thus possible to reduce the input vector from N dimensions to M dimensions

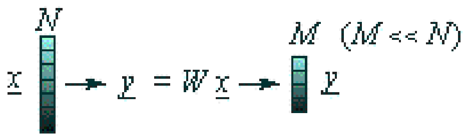

Figure 1:PCA works by compressing the original input vector into a smaller transformed vector whilst retaining the maximum information content (variance) in the transformed vector

The first requirement for the application of the PCA transform is an analysis of the eigenvectors and eigenvalues of the correlation (or covariance) matrix formed from the original input data. Let the eigenvalues of the covariance matrix be denoted by $\lambda_i \; i = 1 \ldots N$ and arranged such that they are monotonically decreasing i.e. $\lambda_i \geq \lambda_{i+1}$.The cumulative eigenvalue curve c(x) is then defined to be

$$c(x) = \frac{\sum_{i=1}^{x} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$$

The interpretation of this curve is that the value c(x) represents the amount of **information** maintained in the input vectors if we project them onto the subspace spanned by the top x eigenvectors. A feature transformation that, for instance, retains 95 percent of the original information (variance)of the input data can be obtained by selecting the appropriate value for x, such that c(x) = 0.95.

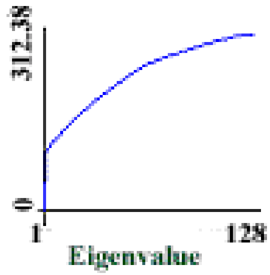

Figure 2 Example of a cumulative eigenvalue curve.

DISCUSSION

There are two important points to b ear in mind when using principal components analysis as a data dimension reduction technique.

PCA retains as much information **in the Shannon sense** as possible in the sub-space projection. This is not necessarily what is required as the important information (the information that helps us to discriminate between up and down future returns for instance) may not be associated with the variance in the input data. For a full discussion of this see the Theoretical Report IFS-TR-007 - *The problem with PCA.*

It is important to appropriately scale the input components before applying PCA. The scale of the components does not need to be identical but should at least be of the same order of magnitude.

PCA has been fully implemented within Amber and is available as a feature transform component.

Author: Darren Toulson
Revision: v 1.01 4 February, 1997
See ALSO: Theoretical Reports IFS-TR-006,IFS-TR-007, The Amber Reference Guide