Principal Component Analysis

Martin Sewell

Department of Computer Science

University College London

April 2007 (revised August 2008)

Principal component analysis (also known as *principal components analysis*) (PCA) is a technique from statistics for simplifying a data set. It was developed by Pearson (1901) and Hotelling (1933), whilst the best modern reference is Jolliffe (2002). The aim of the method is to reduce the dimensionality of multivariate data whilst preserving as much of the relevant information as possible. It is a form of unsupervised learning in that it relies entirely on the input data itself without reference to the corresponding target data (the criterion to be maximized is the variance).

PCA is a linear transformation that transforms the data to a new coordinate system such that the new set of variables, the principal components, are linear functions of the original variables, are uncorrelated, and the greatest variance by any projection of the data comes to lie on the first coordinate, the second greatest variance on the second coordinate, and so on. In practice, this is achieved by computing the covariance matrix for the full data set. Next, the eigenvectors and eigenvalues of the covariance matrix are computed, and sorted according to decreasing eigenvalue. Note that PCA's bias is not always appropriate; features with low variance might actually have high predictive relevance, it depends on the application.

References

HOTELLING, Harold, 1933. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, **24**(6 & 7), 417–441 & 498–520.

JOLLIFFE, I. T., 2002. *Principal Component Analysis*. Second ed. Springer Series in Statistics. New York: Springer-Verlag New York.

PEARSON, Karl, 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine, Series 6*, **2**(11), 559–572.