

The Application of Multivariate Projection Methods to the Analysis of Large-Scale Proteomic Data Sets

J.Campbell

Proteome Sciences plc, Institute of Psychiatry, London, UK

Proteome Sciences plc is a company whose primary goal is the discovery and commercialisation of biomarkers indicating the initiation or progression of diseases, efficacy of treatment and underlying causes of drug toxicity. Many of our proteomic data sets consist of quantitative data collected from SELDI mass spectrometry instruments or 2-D gels. These methods are applied to samples collected from patients with, for example, cancers, neurological or cardiovascular diseases as well as carefully selected controls.

Proteomics data sets tend to have high-dimensionality and thus analysis by univariate methods is hazardous because of the risk of missing interesting correlation structures that may be present in the data. We have been applying multivariate projection methods, namely principal components analysis (PCA) and partial least squares - discriminant analysis (PLS-DA) to the analysis of these data sets to extract further meaningful information from them.

To illustrate this approach, a published data set containing SELDI data derived from serum of patients with and without ovarian cancer has been analysed in depth and the results of this analysis compared with other previously published methods. In addition, some brief examples of how we have applied these methods to new 2-D gel and SELDI mass spectrometry data sets are also given.