Partial Least Squares

Regression by means of projections to latent structures (PLS, also known as partial least squares) is a useful alternative to the linear multiple regression model fitted by least squares if at least one of the following conditions applies:

- the *x*-variables, often known as predictors, are correlated;
- the number of *x*-variables is relatively high compared with the number of observations;
- the *y*-variable(s), often known as response(s), are correlated.

Thus the PLS method is popular in industries that collect correlated predictor data, for example, multivariate calibration in analytical chemistry; spectroscopy in chemometrics; and quantitative structure activity relationships in drug design.

The PLS method extracts orthogonal linear combinations of predictors, known as factors, from the predictor data which explain variance in both the predictors and response(s). In general, a PLS analysis consists of the stages:

1. Extract a given number of factors (more than is likely to be required) from data.

2. Select the number of factors to include in a fitted model by either:

    - analysing information calculated during the process of extracting factors or;
    - calculating a prediction accuracy estimate based on, for example, cross-validation.

3. Fit a PLS model with the selected number of factors by estimating parameter values of the linear regression.

4. Use a fitted model to estimate response values to new predictor data.

At Mark 9 of the NAG C Library, we have the PLS functionality to:

(a) Extract factors by using Wold's method (g02lac);

(b) Extract factors by using singular value decompositions (g02lbc);

(c) Fit a model (g02lcc);

(d) Predict response(s) for new data (g02ldc).

Reference

Wold S (1994) PLS for Multivariate Linear Modeling QSAR: Chemometric Methods in Molecular Design. *Methods and Principles in Medicinal Chemistry* van de Waterbeemd H (Editor) Verlag-Chemie.