

Pittsburgh Brain Activity Interpretation Competition 2006

Methods Description

Partial Least Squares Regression

Project Abstract

Partial least squares is a popular method for soft modeling in industrial applications. Partial least squares (PLS) is a method for constructing predictive models when the factors are many and highly collinear. We think the data in our competition are large and highly collinear. So we use PLS to predict the feature rating. We have tried to predict using the method: Back-Propagation Network and Support Vector Machine. Through comparing the result, we find the PLS is the best.

Introduction

Partial least squares is used to get a general best prediction. This method is used first in our institute, so there are not much experience to apply. We have tried to predict using the method: Back-Propagation Network and Support Vector Machine.

Method

The goal of pls regression is to predict Y from X and to describe their common structure. When Y is a vector and X is full rank, this goal could be accomplished using ordinary multiple regression.

The properties of pls regression can be analyzed from a sketch of the original algorithm. The first step is to create two matrices: $E = X$ and $F = Y$. These matrices are then column centered and normalized (i.e., transformed into Zscores). The sum of squares of these matrices are denoted SSX and SSY .

Before starting the iteration process, the vector u is initialized with random values. (in what follows the symbol $/$ means "to normalize the result of the operation").

Step 1. $w = E^T u$ (estimate X weights).

Step 2. $t = Fw$ (estimate X factor scores).

Step 3. $c = F^T t$ (estimate Y weights).

Step 4. $u = Fc$ (estimate Y scores).

If t has not converged, then go to Step 1, if t has converged, then compute the value of b which is used to predict Y from t as $b = t^T u$, and compute the factor loadings for X as $p = E^T t$. Now subtract (i.e., partial out) the effect of t from both E and F as follows $E = E - tp^T$ and $F = F - btc^T$. The vectors t , u , w , c , and p are then stored in the corresponding matrices, and the scalar b is stored as a diagonal element of B . The sum of squares of X (respectively Y) explained by the latent vector is computed as $p^T p$ (respectively b^2), and the proportion of variance explained is obtained by dividing the explained sum of squares by the corresponding total sum of squares (i.e., SSX and SSY).

If E is a null matrix, then the whole set of latent vectors has been found, otherwise the procedure can be re-iterated from Step 1 on.

Then we use the software supplied by the home page of the N-Way project (www.models.kvl.dk/source/nwaytoolbox/).

In the Matlab programme there are two functions as follow:

```
function [Xfactors,Yfactors,Core,B,ypred,ssx,ssy,reg] = npls(X,Y,Fac,show);
```

Input:

X Array of independent variables
Y Array of dependent variables
Fac Number of factors to compute

Output :

Xfactors Holds the components of the model of X in a cell array.
Yfactors Similar to Xfactors but for Y
Core Core array used for calculating the model of X
B The regression coefficients from which the scores in the Y-space are estimated from the scores in the X-space ($U = TB$);
ypred The predicted values of Y for one to Fac components (array with dimension Fac in the last mode)
ssx Variation explained in the X-space.
 ssx(f+1,1) is the sum-squared residual after first f factors.
 ssx(f+1,2) is the percentage explained by first f factors.
ssy As above for the Y-space
reg Cell array with regression coefficients for raw (preprocessed) X
function [ypred,T,ssX,Xres]=npred(X,Fac,Xfactors,Yfactors,Core,B,show)
ypred the predictions of y

We use the Movie1 to train the model to predict the Movie3.

Results and Discussion

The results are not ideal. There are many other method to use because the time is not enough.and this field is new for us .There are much work for us to do.The prediction is a important research field. Partial least squares(PLS) is a method for constructing predictive models when the factors are many and highly collinear.

References

- [1] Hervé Abdi(2003) Partial Least Squares (PLS) Regression. Encyclopedia of Social Sciences
- [2] Randall D. Tobias, An Introduction to Partial Least Squares Regression
- [3] Bates, D.M. & Watts D.G. (1988). Nonlinear regression analysis and its applications. New York.
- [4] Claus A. Andersson, Rasmus Bro(2000) The N-way Toolbox for MATLAB

Elsevier Science B.V.

[5]Ye Ying , CHEN Chong2Guo , LINXi(2005) Theory and Application of Partial Least Squares Regression , Strait J Prev Med 2005 Vol111 No13

[6] ZHANG Fu-sheng , SHORT-TERM LOAD FORECASTING BASED ON PARTIAL LEAST-SQUARES REGRESSION, Power System Technology