Fra: Eide, Geir Egil [mailto:geir.egil.eide@helse-bergen.no]
Sendt: 3. februar 2016 14:22
Til: 'nsf-medlemmer@math.uio.no'; 'bmms-ext@medisin.uio.no' (bmms-ext@medisin.uio.no); 'bmms-int@medisin.uio.no' (bmms-int@medisin.uio.no)
Emne: [biostat-ext] Bergen biostatistical seminar: Thursday 11 February 2016, 11.00-12.00

Dear colleague,

the next **Bergen biostatistical seminar** will be given by

Prof. II Ragnar Nortvedt, PhD, Program Manager, MedViz (http://www.medviz.uib.no/)

Thursday 11 February 2016, 11.00-12.00

Place: Centre for Clinical Research, Armauer Hansen's House, Haukeland University Hospital, room 337 (Bjørgvin)

Title: Characteristics and applications of Partial Least Squares Regression

Abstract: We are facing the following challenges when we want to predict Y from X:

- Lack of selectivity
- Collinearity
- Lack of knowledge

Partial Least Squares (PLS) regression is a flexible method which handle these challenges in a way that in many cases gives good predictions and increased understanding of a complex dataset of multivariate nature. The method requires little prior knowledge about causal relationships, and is thus frequently also called "soft modelling".

PLS is a so-called *bilinear* regression method, in that the **X** matrix itself is approximated by a model that is the product of *two sets* of linear parameters to be estimated, namely the object scores (**T**) and the variable loadings (**P**) from Principal Component Analysis (PCA). Each factor or principal component (PC) can be expressed by **X = TP'** + the residual matrix **E** of unexplained variance. If the residual matrix is high after the identification of the first PC, we may continue to extract information from the remaining dataset by identifying more PCs. In the PLS regression analysis, the loading vector **P** represents the regression coefficients of **X** on **T** and **Q** represents the regression coefficients of **Y** on **T**. Similarly, we can also write **Y = TQ'** + the residual matrix **F**. The initial model should be based upon a *calibration dataset*. The final model should only include factors that improve the prediction of **Y** in independent *test objects*. Graphical interpretation is also important.

The current presentation will discuss some of the characteristics of PLS and present examples of more or less useful applications of PLS regression.

The seminar is open for everyone interested. After the seminar there will be a light lunch.

Welcome!

Geir Egil Eide

PS: Reserve also your time 6 April at 11.00 for the next Bbiss on statistical methods for big data.