

Re: Output from Model Comparison based on prediction Error

Inge Svein Helland <ingeh@math.uio.no>

Thu 11/24/2016 10:49 AM

To: Raju Rimal <raju.rimal@nmbu.no>;

Cc: Solve Sæbø <solve.sabo@nmbu.no>; Trygve Almøy <trygve.almoy@nmbu.no>;

Dear Raju,

Thank you for your files. You have done a great job! And the results seem to be very encouraging for BayesPLS.

The question is how we should present all this information in the paper. I have some suggestions, but you should also make up your own opinions and discuss these things with Solve and Trygve.

1. I think we only should present a couple of these design plots, and choose only plots with respect to test sets.
2. In presenting the plots, I think we should drop the q , and just say in the text that $q=p$.
3. In the further discussion, I think we should drop the designs where ols is best in the test set. These designs are cases where you should not use any biased regression method at all. As far as I can see, these are designs 2, 6, 10, 14, 18, 22, 26 and 30.
3. To make the differences between methods more visible, might it be an idea to plot something like $[PE(k, \text{method}) - PE(k, \text{PLS})] / PE(k, \text{PLS})$ for $k=1, 2, \dots, 10$?
Here PE is the prediction error that you have plotted, k is the number of components, and the methods that are of main interest is BayesPLS and the envelope. The y-axis of these plots should be rather short, and we could have many designs under each other, identified only by number of design.
4. It may also be an idea to make a table of the mean over k (1,...10) of $PE(k, \text{method})$ for the designs of interest and the methods PLS, BayesPLS, PCR and envelope. Then the designs should only be presented by numbers and all the numbers defined in a separate table.

The challenge is to get much relevant information across and at the same time keeping the presentation brief.

Best regards
Inge

PS. You also attached a htm-file. I was not able to open that file on my mac.

From: Raju Rimal <raju.rimal@nmbu.no>

Sent: 22 November 2016 16:54

To: Inge Svein Helland

Cc: Solve Sæbø; Trygve Almøy

Subject: Output from Model Comparison based on prediction Error

Dear Inge,

I have some output of prediction error plot of different model. It seems, the BayePLS is the best interns of prediction error. However, it is very time consuming to fit all Bayes model for each components. I have included the design properties in each plot. The notations are,

n: Number of training samples

p: number of predictor variables

R2: Coefficient of determination

relpos: position of relevant components

gamma: decay factor of eigenvalues. High value represents fast decrease in eigenvalue

q: number of relevant predictors which we have set equals to p

Best Regards,

Raju Rimal

Norwegian University of Life Science

Norway

rajurim@nmbu.no