Re: Using Ridge Regression to disentangle highly correlated explanatory variables

Source: http://sci.tech−archive.net/Archive/sci.stat.math/2008−03/msg00119.html

- *From*: Paige Miller <paige.miller@xxxxxxxxx>

- *Date*: Fri, 14 Mar 2008 12:34:05 −0700 (PDT)

On Mar 13, 2:25 pm, JohnF <jf...@xxxxxxxxxxx> wrote:

> Folks,
>
> Need your advice and any practical solutions.
>
> We recently conducted a retrospective regression analysis where 3
> variables were highly correlated (high VIFs). Decided to use a
> principal components approach to create a factor score for input into
> the regression model, which did it's job at reducing the VIF greatly.
>
> However, the three highly correlated variables were each of great
> interest. A colleague suggested using Ridge Regression to disentangle
> the relative impact of each of the three explanatory variables. This
> did show that one of the three variables was much more impactful.
>
> Now I'm left wondering if this makes sense, given they were so highly
> correlated to begin with. Wouldn't we conclude that they are all
> equally contributing − i..e, the factor loading can be divided in
> terms of relative impact equally among the three variables?
>
> What's your opinion on this type of issue. I need some practical
> advice, point of view, and/or alternate approach to consider.
> Remember that the three variables are each of particular interest, so
> need to somehow cull out their relative impact.
>
> Very much appreciate any and all help. Thanks!
>
> John

You haven't explicitly stated the goal of the regression here. It is implied, as I read between the lines, to be understanding the independent individual effects of the three highly correlated predictors. (If you are trying to obtain a good prediction equation over your data space without trying to understand the individual impacts of your three correlated predictors, that leads to different answers)

I agree with the other commenters that there really isn't a way to use this data to obtain independent individual effects of the three highly correlated predictors. Logically, can't be done (does not depend on the solution algorithm, it is a logical impossibility). I might suggest that if this is a very important problem, you should consider performing a designed orthogonal experiment to get the information you need. I realize that some fields of study, like econometrics, don't lend themselves to designed orthogonal experiments.

I have never heard that Ridge Regression can be used to "to disentangle the relative impact of each of the three explanatory variables". I always thought of it as a biased estimation method which leads to equations that have better precision (and lower mean squared error) than OLS. In any event, the fact that your ridge regression showed one variable to be "more impactful" than the others only means that you found a transformation of your original data where the one predictor appeared to be more impactful. I'm sure there are other transformations (different ridge parameters) that might give different results.

I am also surprised that your use of PCA "did it's job at reducing the VIF greatly". If you do PCA properly, the inputs are now uncorrelated, and the VIFs should be 1 for each score variable.

Bottom line, you can't get where you want to go from your starting point. You can estimate unbiased prediction equations via OLS where the variances of the three highly correlated predictor's coefficients are huge; or you can estimate biased prediction equations via RR or PLS where the variances of the three highly correlated predictor's coefficients are noticeably reduced compared to OLS; but you can't disentangle the effects of the three predictors.

––
Paige Miller
paige\dot\miller \at\ kodak\dot\com
.