On Reduced Rank Regression

Notes by Parker Knight

January 21, 2022

1 Introduction

Consider the multivariate linear model

$$\mathbb{Y} = \mathbb{X}B + E$$

where $\mathbb{Y}, E \in \mathbb{R}^{n \times q}$, $\mathbb{X} \in \mathbb{R}^{n \times p}$ and $B \in \mathbb{R}^{p \times q}$. In other words, we have n samples, q distinct outcomes, and p features or covariates. For now, we assume $n \geq p$ and $\mathbb{X}$ is full rank. We can estimate B via least squares:

$$\begin{aligned}
\hat{B}_{OLS} &= \arg\min_B \|\mathbb{Y} - \mathbb{X}B\|_F^2 \\
&= (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{Y}
\end{aligned}$$

Note that this is equivalent to running q separate univariate regression, and combining the coefficient estimates into a matrix. This fails to share information across each of the q regression problems.

2 Reduced Rank Regression

One possible hypothesis is that the signal in B is driven by a small number of latent factors; i.e., B is low rank. This motivates *reduced rank regression* [1], in which we estimate B by solving the following:

$$\hat{B}_{RRR}(k) = \arg\min_{B:\text{rank}(B) \leq k} \|\mathbb{Y} - \mathbb{X}B\|_F^2$$

We will see that a convenient, closed form solution exists.
Notice

$$\begin{aligned}
\|\mathbb{Y} - \mathbb{X}B\|_F^2 &= \|(\mathbb{Y} - \mathbb{X}\hat{B}_{OLS}) + (\mathbb{X}\hat{B}_{OLS} - \mathbb{X}B)\|_F^2 \\
&= \|(\mathbb{Y} - \mathbb{X}\hat{B}_{OLS})\|_F^2 + \|(\mathbb{X}\hat{B}_{OLS} - \mathbb{X}B)\|_F^2
\end{aligned}$$

The second equality holds because $\mathbb{X}\hat{B}_{OLS}$ is the orthogonal projection of $\mathbb{Y}$ onto the column space of $\mathbb{X}$, and so $\mathbb{Y} - \mathbb{X}\hat{B}_{OLS}$ is orthogonal to this column space. Since $\mathbb{X}\hat{B}_{OLS} - \mathbb{X}B$ clearly resides in the column space of $\mathbb{X}$, the cross terms are zero and we have equality.

Since $\mathbb{Y} - \mathbb{X}\hat{B}_{OLS}$ does not include B, it follows that we can write

$$\begin{aligned}
\hat{B}_{RRR}(k) &= \underset{B:\text{rank}(B)\leq k}{\arg\min} \|\mathbb{X}\hat{B}_{OLS} - \mathbb{X}B\|_F^2 \\
&= \underset{B:\text{rank}(B)\leq k}{\arg\min} \|\mathbb{X}(\hat{B}_{OLS} - B)\|_F^2 \\
&= \underset{B:\text{rank}(B)\leq k}{\arg\min} \text{tr}\left[(\hat{B}_{OLS} - B)^T\mathbb{X}^T\mathbb{X}(\hat{B}_{OLS} - B)\right] \\
&= \underset{B:\text{rank}(B)\leq k}{\arg\min} \text{tr}\left[(\hat{B}_{OLS} - B)^T\hat{\Sigma}(\hat{B}_{OLS} - B)\right] \\
&= \underset{B:\text{rank}B\leq k}{\arg\min} \|\hat{B}_{OLS} - B\|_{\hat{\Sigma}}^2
\end{aligned}$$

This is precisely the rank k Generalized Matrix Decomposition (GMD) of $\hat{B}_{OLS}$ with respect to $\hat{\Sigma} = \mathbb{X}^T\mathbb{X}$; see [2, 3]. Thus we can compute $\hat{B}_{RRR}(k)$ easily.

3 Generalizations

Can we generalize this notion of a GMD-based reduced rank regression? For instance, consider the more general problem:

$$\hat{\mathcal{B}}(k) = \underset{B:\text{rank}(B)\leq k}{\arg\min} \|\hat{\Theta} - B\|_\Psi^2 + P_\lambda(B)$$

where $\hat{\Theta} \in \mathbb{R}^{p\times q}$ is any collection of regression coefficients for the q outcomes[1], Ψ is any matrix of similarity/correlation between features, and $P_\lambda(.)$ is a penalty function with tuning parameter $\lambda > 0$. This problem is a penalized, generalized matrix decomposition [4] on a matrix of regression coefficients.

One application that I'm interested in: what if $\hat{\Theta}$ is a matrix of SNP-phenotype associations from a GWAS, and Ψ is an LD matrix?

References

[1] A. J. Izenman, "Reduced-rank regression for the multivariate linear model," vol. 5, no. 2, pp. 248–264.

[2] G. I. Allen, L. Grosenick, and J. Taylor, "A generalized least-square matrix decomposition," vol. 109, no. 505, pp. 145–159.

[1]Perhaps, for instance, the columns of $\hat{\Theta}$ are computed by a LASSO or Ridge regression.

[3] Y. Wang, A. Shojaie, T. W. Randolph, and J. Ma, "Generalized matrix decomposition regression: Estimation and inference for two-way structured data,"

[4] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," vol. 10, no. 3, pp. 515–534.