HCP data application

Chanwoo Lee, August 27, 2020

1 Parameter tuning

I perform simple simulation to set the range of smooth parameter that we would like to consider. Since HCP has the matrix size 68×68, I set matrix size as 60×60. This simulation is based on R-code "simulation.R" file. The number of sample size is set to be 212 which is the same sample size of HCP. I changed cost value from 0.1 to 10.When cost value is too small or too small, true continuous value y_{true} and fitted $\hat{y}$ does not have similar value. Therefore, I set considered cost parameter in interval $\{1, \ldots, 8\}$.

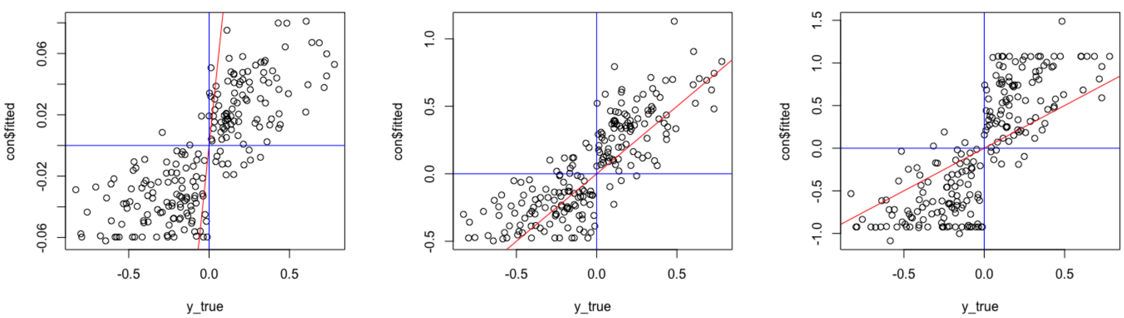

Figure 1: The Figures plot true y and fitted y from different cost when rank is 1. Red lines plot $y = x$ and blue lines represent axis. From left to right, cost is 0.5,4,10 in order.

I set rank from 1 to 20 as considered rank range. From, those hyper parameter ranges, I perform cross validation for HCP. Cross validation procedure is as follows

1. Divide samples into 5 sets such that the number of $y = 1$ and $y = -1$ are the same in each set.
2. For (cost, rank) $\in \{1, \ldots, 8\} \times \{1, \ldots, 20\}$, perform 5 folded cross validation.
3. Chosse the hyperparameter which has the smallest average of miss classification errors on the 5 test set.

I currently requested jobs on the server.

2 Interpretation of the result

In the reference paper, the goal of analysis is to study relationship between brain structural connectivity and visuospatial processing. We can find this relationship from our model.

$$y_i = \text{sign}\left(\langle \boldsymbol{B}, \boldsymbol{X}_i \rangle + b_0\rangle\right),$$

where $\boldsymbol{X}_i$ is an adjacent matrix of i-th subject and y_i is binary label which is decoded as 1 if i-th subject is in high visuospatial processing group and -1 otherwise. We interpret the value of $\boldsymbol{B}$ as measure of contribution to visuospatial processing. To be specific, if $\boldsymbol{B}_{ij}$ value is a large

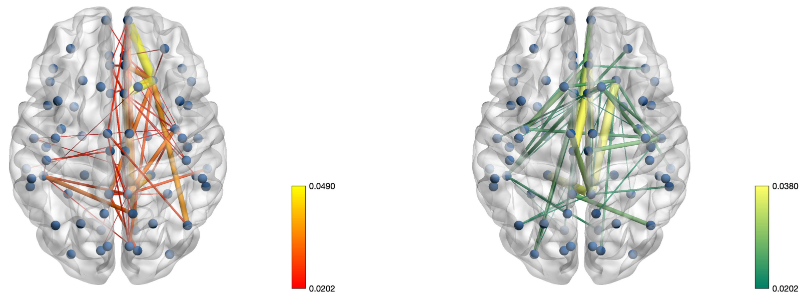

Figure 2: The left figure shows connections that have positive effect on the VSPLOT and the right figure plots connections that have negative effect on the VSPLOT. Only the connections that have greater than 0.02 magnitude are plotted

positive value, then connection between $i-th$ and $j-th$ brain nodes has a strong positive effect on visuospatial processing. On the other hand, if $\boldsymbol{B}_{ij}$ value is a large negative value, the connection has a strong negative effect on the capacity. Therefore, by plotting the connections on the brain which is greater than certain positive threshold, we can visualize connections of brain nodes which has positive effect. By the same way, we can plot negative effect connection between nodes. The following plot is the output of random trial when $r = 5$ and cost =1. Based on the fact that visuospatial processing is linked to posterial cortical function, the current figure does not make sense.

Remark 1. For interpretation, I used linear kernel because we have direct interpretation on the coefficient $\boldsymbol{B}$. I am not sure how to relate brain connection and VSPLOT from output of our algorithm when other kernel method used

Remark 2. Since current algorithm give us output $\boldsymbol{P}_{\text{row}}, \boldsymbol{P}_{\text{col}}, \boldsymbol{\alpha}$. Based on this I calculated $\boldsymbol{B}$ from the following procedure

1. Calculate coefficient $\boldsymbol{B}$ that might have the rank greater than prespecificed rank as follows

$$\boldsymbol{B} = \sum_{i=1}^{n} \alpha_i y_i \left(\boldsymbol{P}_{\text{row}} \boldsymbol{P}_{\text{row}}^T \boldsymbol{X}_i + \boldsymbol{P}_{\text{row}} \boldsymbol{P}_{\text{row}}^T \boldsymbol{X}_i \right)$$

2. Since $\boldsymbol{X}_i$ is symmetric matrix make $\boldsymbol{B}$ symmetric as $\boldsymbol{B} = \frac{\boldsymbol{B}+\boldsymbol{B}^T}{2}$. This symmetrization does not change the output because of symmetricity of $\boldsymbol{X}$

3. Find the best rank r-approximation of $\boldsymbol{B}$.