Graphic Lasso: Gene expression data analysis

Jiaxin Hu

January 17, 2021

1 Settings

Let $\rho = 2000, r = 2$. The corresponding model is

$$\Omega^k = \Theta_0 + u_{k1}\Theta_1 + u_{k2}\Theta_2,$$

where $\boldsymbol{U} = [\![u_{ij}]\!] \in \mathbb{R}^{K\times r}$ has non-overlapped columns. The Matlab code converges with tolerance 10^{-3} in 88 iterations.

2 Membership Results

Based on $\hat{\boldsymbol{U}}$, we obtain the membership result for 13 brain tissues. See Table 1.

$A_0 : u_{k1} = u_{k2} = 0$	Anterior cingulate cortex
$A_1 : u_{k1} \neq 0, u_{k2} = 0$	Cerebellum, Spinal cord, Amygdala, Hippocampus, Substantia nigra Frontal Cortex, Cerebellar Hemisphere, Hypothalamus
$A_2 : u_{k1} = 0, u_{k2} \neq 0$	Caudate (basal ganglia), Nucleus accumbens (basal ganglia), Putamen (basal ganglia), Cortex

Table 1: Membership result for 13 brain tissues.

The membership result makes sense. First, Anterior cingulate cortex (ACC) is the frontal part of cingulate cortex while other tissues belong to cerebellum, basal ganalia and cerebrum. Thus, ACC has different functions with other tissues and has a different gene expression performance. Second, except the cortex, the tissues in group A_1 are all nearby the cerebellum, and the tissues in group A_2 are all in basal ganglia. Since both cerebellum and basal ganalia receive input from and send output to the cerebral cortex, it makes sense that cortex appear in both A_1 and A_2.

3 Gene Correlation Results

3.1 $\hat{\Theta}_1$

In $\hat{\Theta}_1$, there exist 8 nonzero off-diagonal elements. See Table 2 for gene pairs with top 5 correlation.

Gene	Function	Gene	Function	$\Theta_{1,ij}$
PENK	Height	PDE10A	Blood urea nitrogen levels, height	-0.14522
RXFP1	Blood metabolite ratio	PDYN	Blood protein levels	-0.11291
TBR1	Education attainment	SLC17A6	Creatine kinase levels, blood pressure	-0.083288
DAO	Platelet crit	SLC17A6	Creatine kinase levels, blood pressure	-0.043183
GPR6	Platelet distribution width	SCN4B	Blood protein levels	-0.022352

Table 2: Non zero correlation indicates in $\hat{\Theta}_1$. Here only records top 5 correlations.

The correlation result makes sense. First, both of two genes with the top 1 correlation are significant to Height in GWAS. Second, the other genes with correlations are both related with blood. Last, recent research also suggests that education status is inversely correlated with the blood pressure, which implies the third correlation is reasonable.

3.2 $\hat{\Theta}_2$

In $\hat{\Theta}_2$, there exist 57 nonzero off-diagonal elements. See Table 3 for gene pairs with top 5 correlation.

Gene	Function	Gene	Function	$\Theta_{2,ij}$
DLK1	Platelet crit	SLC6A3	red cell distribution width	-0.19933
CCKAR	Varicose veins of lower extremity	SLC6A3	red cell distribution width	-0.1797
DRD2	Smoking initiation	SLC6A3	red cell distribution width	-0.13096
SV2C	Mean platelet volume	DLK1	Platelet crit	-0.11811
SST	Blood urea nitrogen levels	CRHBP	Mean platelet volume	0.087175

Table 3: Non zero correlation indicates in $\hat{\Theta}_2$. Here only records top 5 correlations.

The correlation result makes sense. First, except DRD2, other genes are related with blood. Second, recent research implies that smoking initiation has inverse relationship with blood pressure which may related to the performance of red cell.

4 Others

The algorithm is sensitive to the tuning parameter ρ. We may try another ρ for better results. Also, we should try other r.