"Sparce inverse covarience estimation with the graphical Lasso"

Paper Report

BY

Mohsen Nabian

This paper considers the problem of estimating sparse graphs by a Lasso penalty applied to the inverse Covarience matrix.

Suppose we have N multivarable normal observation of dimension p : $[X]_{N\times P}$ = observation matrix

The problem is to estimate True covarience $\sum_{P\times P}$

using observation $[X]_{N\times P}$.

Given $\Theta = \Sigma^{-1}$, The whole idea is to maximize log-liklihood:

maximization with respect to Teta

$$\max \{ \log \det \Theta - tr(S\Theta) - \rho \|\Theta\|_1 \}$$

where tr is trace function and Ro is regularization parameter regularization also impose sparcity in covarience matrix.

$\boxed{\text{Algorithm}}$:

① calculate $[S]_{P\times P}$ which is emperical Covarience of $[X]_{N\times P}$

② $W_{P\times P} = S_{P\times P} + \rho I_{P\times P}$ which ρ is Const, lets say $\rho = 0.01$

(3) For each $j = 1, 2, \dots, p, 1, 2, \dots p, \dots$ solve the following Lasso problem which takes as input the inner product W_{11} and S_{12}.

$$\min_{\beta} \left\{ \frac{1}{2} \| W_{11}^{1/2} \beta - b \|^2 + \rho \|\beta\|_1 \right\} \qquad (1)$$

where $b = W_{11}^{-1/2} S_{12}$.

we can solve this Lasso problem by Coordinate descent (Friedman 2007):

Letting: $V = W_{11}$ $\qquad u = S_{12}$

$$\hat{\beta}_j \leftarrow S\Big(u_j - \sum_{k \neq j} V_{kj} \hat{\beta}_k\Big) / V_{jj}$$

For $j = 1, 2, \dots p, 1, 2, \dots p$, until convergance.

where S is the soft~~mo~~threshold operator:

$$S(x, t) = \text{sign}(x)(|x| - t)$$

in ③ in each step j, we find a $p-1$ vector solution $\hat{\beta}$. Then we fill in the corresponding row and column of W using $\boxed{w_{12} = W_{11}\hat{\beta}}$

④ Continue until convergance:

The average absolute change in W is less than $t \cdot ave|S^{-diag}|$ where S^{-diag} are the off-diagonal elements of emperical covarience matrix S and t is a fixed threshold $= 0.001$

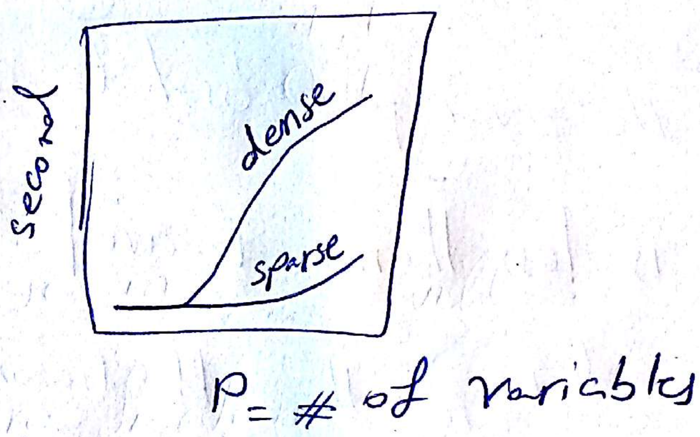
Experimental results:

Graphical lasso is compared with COVSEL program provided by Banerjee. for tow types of dataset

① sparse: $(\Sigma^{-1})_{ii} = 1$ $(\Sigma^{-1})_{i,i-1} = (\Sigma^{-1})_{i-1,i} = 0.5$
zero otherwise

② dense: $(\Sigma^{-1})_{ii} = 2$, $(\Sigma^{-1})_{ii'} = 1$ otherwise results shows Graphical Lasso is about 35 times faster (at Least).

also the number seconds for the graphical Lasso is in two types of problems are depicted:



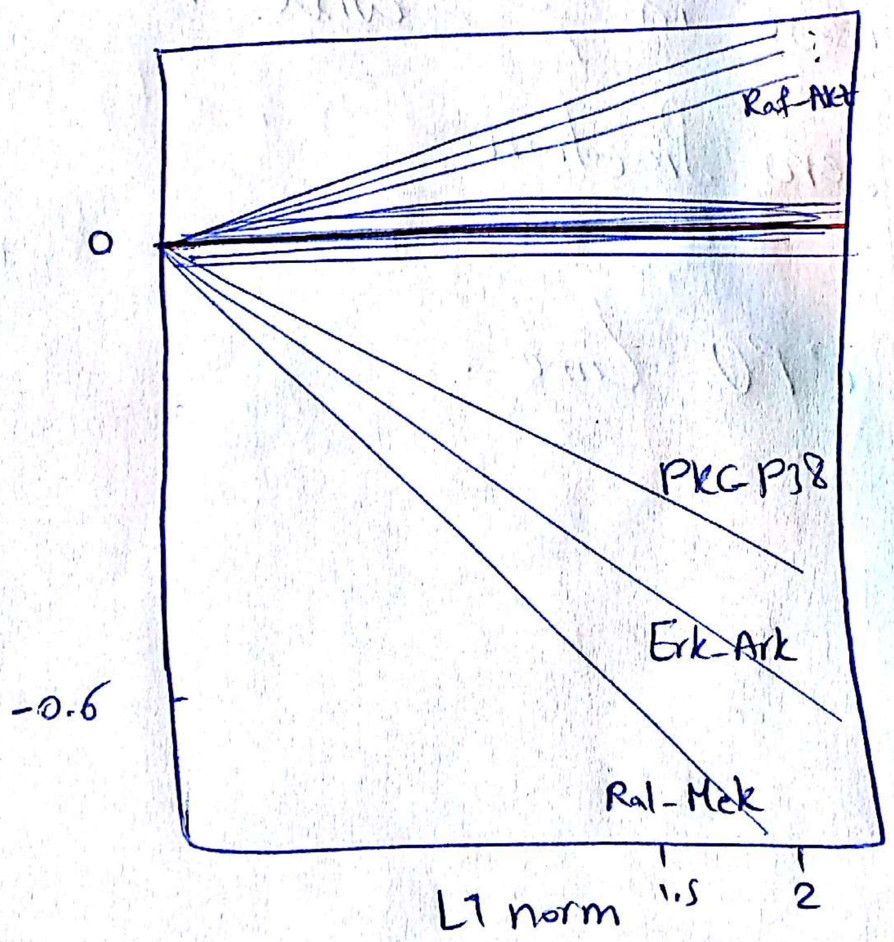
results shows the algo works well for sparse dataset.

Experiment 2 : Analysis of cell signalling data

a data set on p=11 proteins & n = 7466 cells.
Authors fitt a DAG (Sachs 2003)
now we use the data and graphical Lasso to produce same DAG. we use different ρ (Penalty Parameter).

The paper provided the result of 12 different ρ (penalty parameters). for some, which there was good agreement with the DAG.

Also ~~for~~ the pair coefficients of the covarience matrix is provided.



L1 norm $\simeq \rho$ decrease

increase

Another experiment

for a dense senario $p = 400$,
for different regularization params ρ:

ρ	Fraction non-zero	CPU time (sec.)
0.01	.96	26.7
0.03	.62	8.5
0.06	.36	4.1
0.60	.00	0.4

as shown, increase in ρ, will:

① ~~increase~~ decrease the non zero fraction

② decrease the CPU time.