Fused elastic net EPoC

Tobias Abenius

November 28, 2013
Data Set and Method

Use mRNA and copy number aberration (CNA) to build a network using an extension of our EPoC method (Jörnsten et al., 2011; Abenius et al., 2012). Extension with generalized fused LASSO and elastic net generalized fused LASSO which is for this project.

Two different cancers, breast cancer (N=766) and ovarian cancer (N=560). Small subset \((p = 250)\) of all 10321 genes that exist in the intersect gene list of the two cancers.
EPoC summary

The EPoC method first standardize mRNA and CNA, then CNA direct effects on their corresponding mRNA are calculated (univariate regression) and subtracted from mRNA to form the residual mRNA.

This residuals will act as response and CNA will act as predictors.
**Fused Objective**

The objective of the standard fused lasso penalizes differences between consecutive coefficients (Tibshirani et al., 2005):

$$
\beta_{\text{fused}} = \arg \min_{\beta} \| y - X\beta \|^2_2 + \lambda_1 \|eta\|_1 + \lambda_2 \sum_{i=2}^{p} |\beta_{i-1} - \beta_i|
$$

This has been extended to a more general case where all pairwise differences are penalized (Petry et al., 2011):

$$
\beta_{\text{pwfused}} = \arg \min_{\beta} \| y - X\beta \|^2_2 + \lambda_1 \|eta\|_1 + \lambda_2 \sum_{i=1}^{p} \sum_{j=i+1}^{p} |\beta_i - \beta_j|
$$

pairwise fused LASSO part

standard fused LASSO part
Fused Objective

The objective of the standard fused lasso penalizes differences between consecutive coefficients (Tibshirani et al., 2005):

$$\beta_{\text{fused}} = \arg\min_{\beta} \| y - X\beta \|_2^2 + \lambda_1 \| \beta \|_1 + \lambda_2 \sum_{i=2}^{p} |\beta_{i-1} - \beta_{i}|$$

This has been extended to a more general case where all pairwise differences are penalized (Petry et al., 2011):

$$\beta_{\text{pwfused}} = \arg\min_{\beta} \| y - X\beta \|_2^2 + \lambda_1 \| \beta \|_1 + \lambda_2 \sum_{i=1}^{p} \sum_{j=i+1}^{p} |\beta_i - \beta_j|$$
This can be even more generalized to (Ye and Xie, 2010)

\[ \beta = \arg \min_{\beta} \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|L\beta\|_1, \]

where matrix \( L \in \{-1, 0, 1\}^{m \times p} \) tell which absolute differences between \( \beta \) coefficients to penalize.
For standard fused lasso $\mathbf{L}$ becomes a matrix with ones on the diagonal and $-1$ on the superdiagonal.

\[
\mathbf{L} = \begin{pmatrix}
\beta_1 & \beta_2 & \cdots & \cdots & \beta_p \\
1 & -1 & 0 & \cdots & 0 \\
0 & 1 & -1 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & 1 & -1
\end{pmatrix}
\]
For pairwise fused lasso $L$ contains $\binom{p}{2}$ rows where a row contains one 1 and one $-1$ in the places corresponding to the coefficient combination for this row. Here is an example with $p = 4$:

$$L = \begin{pmatrix}
\beta_1 & \beta_2 & \beta_3 & \beta_4 \\
1 & -1 & 0 & 0 \\
1 & 0 & -1 & 0 \\
0 & 1 & -1 & 0 \\
1 & 0 & 0 & -1 \\
0 & 1 & 0 & -1 \\
0 & 0 & 1 & -1 \\
\end{pmatrix}$$
Ye and Xie (2010) introduced a method to solve this generalized fused lasso with this $L$:

\[
\begin{align*}
\text{repeat} \\
\beta^{k+1} &\leftarrow \arg \min_{\beta} V(\beta) + \langle u^k, \beta - a^k \rangle + \langle v^k, L\beta - b^k \rangle + \\
&+ \frac{\mu_1}{2} \| \beta - a^k \|^2_2 + \frac{\mu_2}{2} \| L\beta - b^k \|^2_2 \\
a^{k+1} &\leftarrow T_{\frac{\lambda_1}{\mu_1}} \left( \beta^{k+1} + \frac{u^k}{\mu_1} \right) \\
b^{k+1} &\leftarrow T_{\frac{\lambda_2}{\mu_2}} \left( L\beta^{k+1} + \frac{v^k}{\mu_2} \right) \\
u^{k+1} &\leftarrow u^k + \delta_1 (\beta^{k+1} - a^{k+1}) \\
v^{k+1} &\leftarrow v^k + \delta_2 (L\beta^{k+1} - b^{k+1}) \\
\text{until convergence}
\end{align*}
\]

where $T_{\lambda}$ is the soft threshold operator with parameter $\lambda$, $\mu_1 = 1$, $\mu_2 = 1$, $\delta_1 = \mu_1^{-1}$ and $\delta_2 = \mu_2^{-1}$. 
Elastic net

Elastic net: \[ \arg\min_{\beta} \| y - X\beta \|_2^2 + \lambda_{\text{elastic}} \| \beta \|_2^2 + \lambda_{\text{LASSO}} \| \beta \|_1 \]

It is fairly easy to show (see exercise 2.9) that elastic net can be solved by a LASSO solver. Instead of using \( X^T X \) use

\[ G = (1 - \gamma)X^T X + \gamma I_p \]

where

\[ \gamma = \frac{\lambda_{\text{elastic}}}{1 + \lambda_{\text{elastic}}} \cdot \]

I implemented fused elastic net in C as an R package using this extension of the split bregman algorithm.
Results

There seems to be an interplay of different penalty parameters. Up, no elastic penalty, down: $\lambda_{Elastic} = 10$. Horizontal: different $\lambda_{LASSO}$; Vertical: different $\lambda_{Fused}$.

```
xtable(fusedlinks)
```

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1478</td>
<td>1567</td>
<td>1227</td>
<td>44</td>
<td>20</td>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1621</td>
<td>1839</td>
<td>1468</td>
<td>58</td>
<td>29</td>
<td>8</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

```
xtable(fusedelasticlinks)
```

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4148</td>
<td>4640</td>
<td>4352</td>
<td>846</td>
<td>232</td>
<td>62</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>9650</td>
<td>9006</td>
<td>5845</td>
<td>682</td>
<td>182</td>
<td>60</td>
<td>16</td>
<td>0</td>
</tr>
</tbody>
</table>
Results

\texttt{xtable(t(ns1))}

\begin{tabular}{cccccccc}
  & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
 1 & 8553 & 8338 & 5762 & 1129 & 361 & 73 & 21 & 15 \\
 2 & 8553 & 8357 & 5780 & 1137 & 369 & 77 & 29 & 12 \\
\end{tabular}

\texttt{xtable(t(ns2))}

\begin{tabular}{cccccccc}
  & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
 1 & 7973 & 7804 & 6277 & 1320 & 404 & 87 & 44 & 21 \\
 2 & 7970 & 7798 & 6283 & 1315 & 405 & 76 & 31 & 12 \\
\end{tabular}
Claim to verify

Elastic net adds an L2 penalty term and claims that it will not behave as LASSO where of two correlated predictors one can be active and the other not, somewhat randomly. I decided to try to verify this claim.
Results

For each pair of genes (predictors) very correlated ($\rho > 0.9$), how often are exactly one of them in the active set, in all the ($p = 250$) $\times$ ($K = 2$) $\times$ ($\#\lambda_1 = 8$) different regressions?

fused epoc: 9.5179%  elastic fused: 6.4152%

How often are both of them zero?

fused epoc: 65.8906%  elastic fused: 58.2321%

How often are none of them zero?

fused epoc: 24.5915%  elastic fused: 35.3527%
Figure: In blue RSS without elastic and in red with elastic.
References


