Network perturbation analysis of omics data for complex diseases using convex optimization

Luxembourg Center for Systems Biomedicine

Nikos Vlassis & Enrico Glaab

Luxembourg Centre for Systems Biomedicine, University of Luxembourg

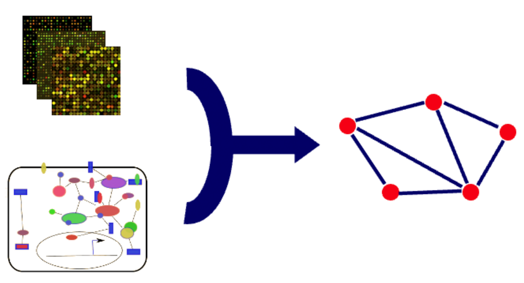
Introduction

Complex diseases like neurodegenerative or cancer disorders are characterized by deregulations in multiple genes and proteins. Previous research[1,2] has shown that neighboring genes in a molecular network tend to undergo coordinated expression changes. We describe an approach that allows identifying such **jointly differentially expressed** genes from input expression data and **a graph encoding pairwise functional associations** between genes (such as protein interactions). We cast this as a **feature selection** problem in penalized two-class (cases vs. controls) classification, and we propose a novel **pairwise elastic net** (PEN) penalty that favors the selection of discriminative genes according to their connectedness in the interaction graph. Experiments on large-scale gene expression data for Parkinson's disease demonstrate marked improvements in feature grouping over competitive methods

Supervised feature selection and grouping

Inputs:

1. Gene expression data $\{\mathbf{x}_i, y_i\}_{i=1}^n$ with features $\mathbf{x}_i \in \mathbb{R}^p$ and class labels $y_i = \pm 1$
2. A graph encoding **pairwise** functional associations between genes (such as protein interactions)



Output: A discriminative set of **connected** genes in the graph.

Penalized logistic regression

Find **weights** $\mathbf{w} \in \mathbb{R}^p$ and $\nu \in \mathbb{R}$ that solve the program

$$\min_{\mathbf{w},\nu} \; f(\mathbf{w},\nu) + \lambda\,\Omega(\mathbf{w}),$$

where $f(\mathbf{w},\nu)$ is the (smooth and convex) expected **logistic loss**

$$f(\mathbf{w},\nu) = \frac{1}{n}\sum_{i=1}^{n} \log\left(1+\exp(-y_i(\mathbf{w}^\intercal \mathbf{x}_i + \nu))\right),$$

and $\Omega(\mathbf{w})$ is a **penalty** term that regularizes $\mathbf{w}$.

Some penalties

$\Omega(\mathbf{w}) = \lVert\mathbf{w}\rVert_2^2$	**Ridge** (Hoerl and Kennard, 1970) grouping but no sparsity
$\Omega(\mathbf{w}) = \lVert\mathbf{w}\rVert_1$	**Lasso** (Tibshirani, 1996) sparsity but no grouping
$\Omega(\mathbf{w}) = \lVert\mathbf{w}\rVert_2^2 + \alpha\lVert\mathbf{w}\rVert_1$	**Elastic Net** (Zou and Hastie, 2005) cannot capture local structure
$\Omega(\mathbf{w}) = \sum_{c\in\mathcal{C}} \alpha_c \lVert\mathbf{w}_c\rVert_2$	**Group Lasso** (Turlach et al., 2005) assumes non-overlapping groups
$\Omega(\mathbf{w}) = \mathbf{w}^\intercal\mathbf{K}\,\mathbf{w}$ (with $\mathbf{K}$ psd)	**graph kernel** (Rapaport et al., 2007) weight signs can introduce bias
$\Omega(\mathbf{w}) = \sum_{i<j} \max(\lvert w_i\rvert, \lvert w_j\rvert)$	**OSCAR** (Bondell and Reich, 2008) large weights can introduce bias

How can we capture graph connectedness of features?

Penalize the differences between **absolute** values of neighboring weights:

$$\Omega(\mathbf{w}) = \sum_{i=1}^{p}\Big[\sum_{j=1}^{p} A_{ij}\lvert w_i\rvert - \sum_{j=1}^{p} A_{ij}\lvert w_j\rvert\Big]^2 + 2\Delta\lVert\mathbf{w}\rVert_1^2,$$

where $\mathbf{A}$ is the **adjacency** matrix of the feature graph, and Δ its maximum **degree**, respectively. This reads:

$$\Omega(\mathbf{w}) = \lvert\mathbf{w}\rvert^\intercal\,\mathbf{P}\,\lvert\mathbf{w}\rvert \qquad \text{where} \qquad \mathbf{P} = \mathbf{L}^2 + 2\Delta\mathbf{J}$$

and where $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the graph **Laplacian**, with $\mathbf{D} = \mathrm{diag}(\mathbf{A}\mathbf{1})$, and $\lvert\mathbf{w}\rvert = (\lvert w_1\rvert, \ldots, \lvert w_p\rvert)$ and $\mathbf{J} = \mathbf{1}\mathbf{1}^\intercal$.

The Pairwise Elastic Net

Our penalty is an instance of the **pairwise elastic net** (PEN) (Lorbert et al., 2010)

$$\Omega(\mathbf{w}) = \lvert\mathbf{w}\rvert^\intercal\,\mathbf{P}\,\lvert\mathbf{w}\rvert,$$

where $\mathbf{P}$ is a $p \times p$ symmetric matrix.

Theorem (Lorbert et al. (2010))

*$\Omega(\mathbf{w})$ is **convex** in $\mathbf{w}$ if and only if $\mathbf{P}$ is **positive semidefinite** and **nonnegative** ($P_{ij} \geq 0\ \forall i,j$).*

In our case, $\mathbf{P} = (\mathbf{I}-\mathbf{A})^2 + 2\mathbf{J}$ is symmetric positive semidefinite, and $\mathbf{P} = \mathbf{D}^2 + \mathbf{A}^2 + (2\Delta\mathbf{J} - \mathbf{D}\mathbf{A} - \mathbf{A}\mathbf{D}) \geq 0$.

Hence our penalty $\Omega(\mathbf{w})$ is convex in $\mathbf{w}$.

Optimization

It is easy to verify that (where '$\succeq$' means entry-wise '$\geq$')

$$\lvert\mathbf{w}\rvert^\intercal\,\mathbf{P}\,\lvert\mathbf{w}\rvert = \min_{\mathbf{u}\succeq\lvert\mathbf{w}\rvert} \mathbf{u}^\intercal\mathbf{P}\mathbf{u}$$

Moreover, the convex set $\mathbf{u} \succeq \lvert\mathbf{w}\rvert$ is equivalent to

$$\{\mathbf{u} = \mathbf{a}+\mathbf{b}, \quad \mathbf{w} = \mathbf{a}-\mathbf{b}, \quad \mathbf{a},\mathbf{b} \in \mathbb{R}_+^p\}$$

PEN as a smooth convex program

$$\min_{\mathbf{a},\mathbf{b},\nu} \; f(\mathbf{a}-\mathbf{b},\nu) + \lambda\,(\mathbf{a}+\mathbf{b})^\intercal\,\mathbf{P}\,(\mathbf{a}+\mathbf{b})$$
$$\text{s.t.} \quad \mathbf{a},\mathbf{b} \in \mathbb{R}_+^p$$

We used the **TFOCS** first-order conic solver (Becker et al., 2011).

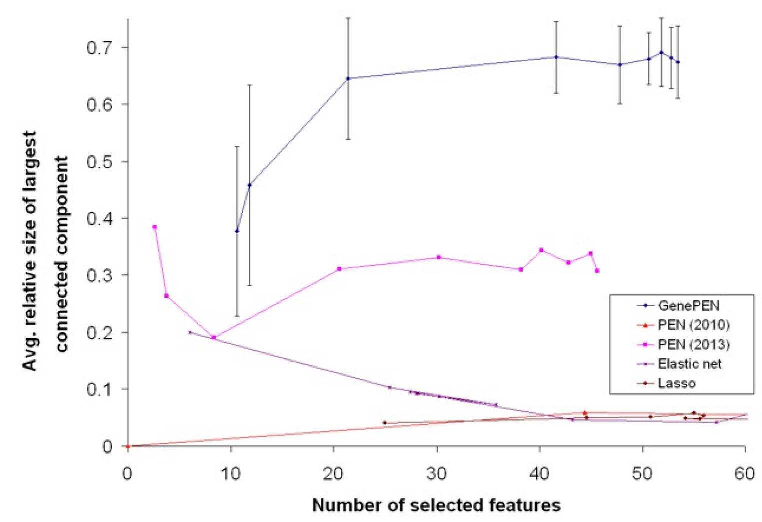
Application: Parkinson's disease gene expression data

We used a publicly available large-scale gene expression dataset (Zhang et al., 2005). This involves $n = 93$ *post mortem* brain samples from **Parkinson's disease** patients (43/93) against unaffected controls (50/93).

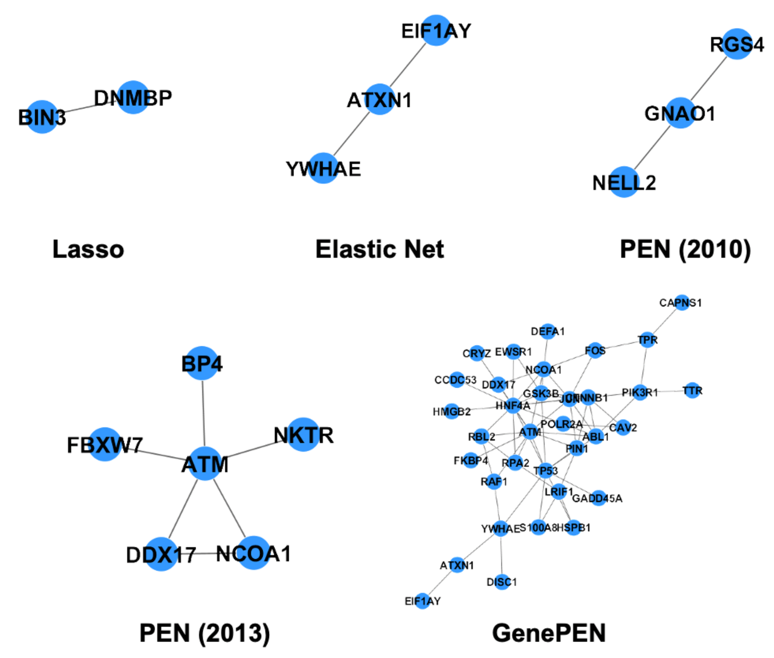
As feature graph, we used a publicly available human genome-scale **protein-protein interaction network** containing 10,042 proteins and 80,543 interactions. We picked up a connected component of $p = 561$ features/genes that are all present in the microarray data.

We tried several penalties $\Omega(\mathbf{w})$. They all exhibited similar cross-validation errors. Hence, our main interest was their **feature grouping** behavior.

Results (1): Relative size of largest connected component



Results (2): Grouping of ~50 features



Results (3): Biological relevance of discovered features

Our method predicts the relevance of **DNM1L** (dynamin-1-like protein), which encodes a protein involved in mitochondrial division. DNM1L is expressed in the brain and loss of DNM1L results in increased oxidative damage in mitochondria, impaired respiration and neurodegeneration in neurons after cell division.

It also predicts **PDGFRB** (platelet-derived growth factor receptor beta), a gene encoding a receptor protein that in a rat model of PD displayed significant changes after neural grafting, and in which mutations have been associated with basal ganglia calcification previously.

All penalties predict **FUS** (fused in sarcoma), which encodes a multifunctional protein involved in cellular processes including regulation of gene expression, maintenance of genomic integrity and mRNA/microRNA processing. Mutations in FUS and aggregations of the corresponding protein have been associated with neurodegeneration.

Conclusions

- We studied biological network deregulation as a **graph-regularized** classification problem.
- We proposed a new penalty for **sparse feature selection and grouping** on a graph, that is based on PEN[4].
- We cast the statistical problem as a **smooth convex program** and solved it with the first-order conic solver TFOCS[6].
- The proposed penalty outperforms other tested penalties for this task[3,4,5].
- Ongoing work involves (i) computing the whole regularization path and (ii) applications in neuroimaging.

References

1. Ideker, T. and Sharan, R. *Genome research* 18(4), 644–652 (2008).
2. Rapaport et al. *BMC Bioinformatics*, 8(1), 35 (2007).
3. Hastie T, Tibshirani R & Friedman J. *The Elements of Statistical Learning*, Springer, 2nd ed. (2009).
4. Lorbert A, Eis D, Kostina V, Blei D.M. & Ramadge PJ. *Proc. AISTATS* (2010).
5. Lorbert A & Ramadge PJ. *Proc. ICASSP* (2013).
6. Becker SR, Candès EJ, Grant MC. *Math Prog Comp* 3(3):165–218 (2011).
7. Zhang Y, James M, Middleton FA & Davis RL. *American Journal of Medical Genetics Part B*, 137(1):5–16 (2005).
8. Bondell, H. D. and Reich, B. J. *Biometrics*, 64(1):115–123 (2008).
9. Hoerl, A. E. and Kennard, R. W. *Technometrics*, 12(1):55–67 (1970).
10. Tibshirani, R. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288 (1996).
11. Turlach, B. A., Venables, W. N., and Wright, S. J. *Technometrics*, 47(3):349–363 (2005).
12. Zou, H. and Hastie, T. *Journal of the Royal Statistical Society: Series B*, 67(2):301–320 (2005).

Luxembourg National Research Fund

Acknowledgments

Alex Lorbert, Martin Jaggi, Rudi Balling

uni.lu UNIVERSITÉ DU LUXEMBOURG