$x \in R^n$ is multivariate Gaussian.

$$x_i \perp x_j \mid x_{-ij} \iff Q_{ij} = 0$$

(Q is precision matrix $Q = \Sigma^{-1}$)

$$\operatorname{corr}(x_i, x_j \mid x_{-ij}) = -\frac{Q_{ij}}{\sqrt{Q_{ii} Q_{jj}}}$$ conditional

$$\operatorname{corr}(x_i, x_j) = \Sigma_{ij} \Big/ \sqrt{\Sigma_{ii} \Sigma_{jj}}$$ marginal

- Estimate Q by maximizing log-likelihood

$$\log \det Q - \operatorname{tr}(SQ)$$

(S: the empirical covariance matrix)

- To enforce sparcity, introduce penalty term.

$$\log \det Q - \operatorname{tr}(SQ) - \lambda \|Q\|_1$$

Existing Solution:

Δ. (Yuan and Lin): Interior point method.

Δ. Banerjee et. al: estimate covariance matrix Σ, insted of its inverse Q

- Frideman: Graphical Lasso

Solve Lasso problem on each variable separately until convergence.

Lasso Problem:

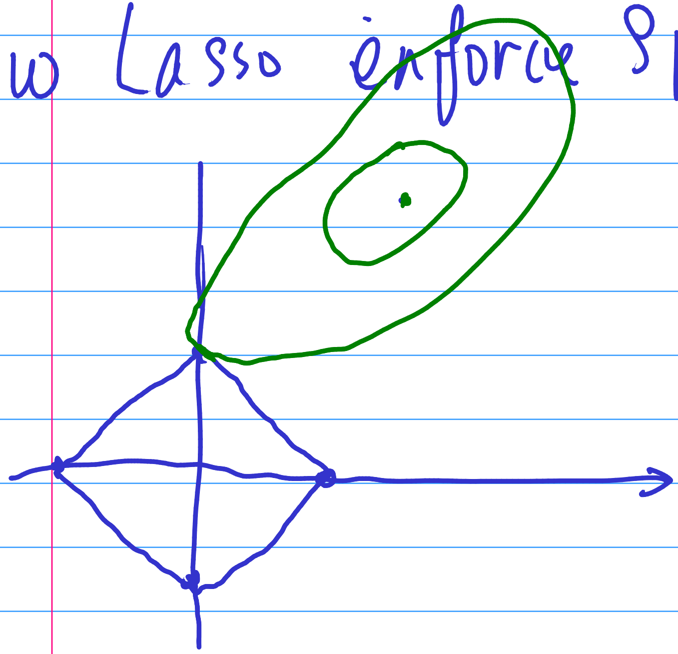
$$\hat{\beta} = \arg\min \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} x_{ij} \beta_j\right)^2$$

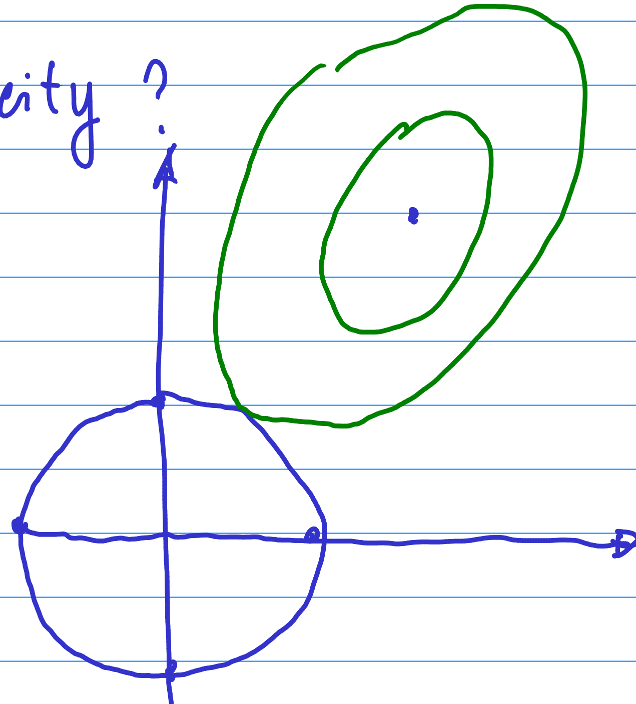
$$\text{subject to } \sum_{j=1}^{p} |\beta_j| \leq t$$

or equivalently

$$f(\beta) = \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} x_{ij} \beta_j\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

How Lasso enforce Sparsity?





Graphical Lasso : Estimate sparseness of the graph (for multivariate Gaussian), i.e. Model selection.

Elastic Net : Regularization term is linear combination of L_1 and L_2 penalty.

$$P_\alpha(\beta) = (1-\alpha)\,\|\beta\|_{\ell_2}^2 + \alpha\,\|\beta\|_{\ell_1}$$

$$\min\left[\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i^T\beta)^2 + \lambda P_\alpha(\beta)\right]$$

- Elastic Net can be used to estimate graph sparseness just like Lasso.

Extention Lasso to multi-subjects of fMRI

Group Lasso :

X : $N \times p$ observation matrix

Y : $N \times 1$ response variable.

p features are grouped into $\beta_1 \cdots \beta_j \cdots \beta_J$

into J groups

Group lasso try to minimize

$$\| y - \sum_{j=1}^{J} X_j \beta_j \|^2 + \lambda \sum_{j=1}^{J} \| \beta_j \|$$

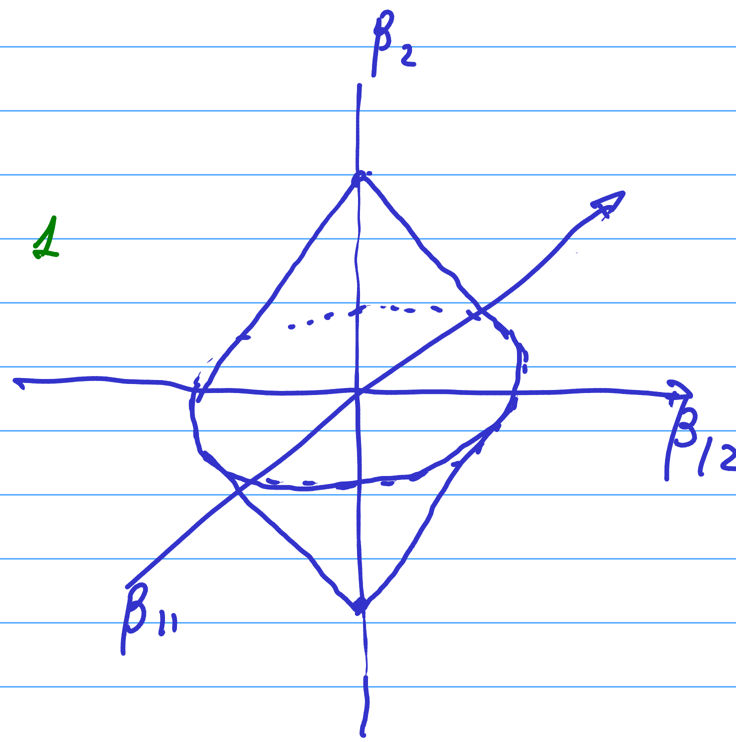
X_j : design matrix for P_j feature in group j

β_j : for group j.

all β in β_j are removed or added into the model together.

β_{11} and β_{12} in same Group 1

β_2 in group 2



Relation to fMRI

Each subject in a group

$$
\underbrace{\begin{bmatrix} y_1^{(1)} \\ \vdots \\ y_N^{(1)} \\ y_1^{(2)} \\ \vdots \\ y_N^{(2)} \end{bmatrix}}_{N \times 1}
=
\underbrace{\left[\begin{array}{ccc|ccc}
x_{11}^{(1)} & \cdots & x_{1p}^{(1)} & & & \\
\vdots & & & & 0 & \\
x_{N1}^{(1)} & \cdots & x_{Np}^{(1)} & & & \\
\hline
& & & x_{11}^{(2)} & \cdots & x_{1p}^{(2)} \\
& 0 & & \vdots & & \vdots \\
& & & x_{N1}^{(2)} & \cdots & x_{Np}^{(2)}
\end{array}\right]}_{N \times P}
\cdot
\underbrace{\begin{bmatrix} \beta_1^{(1)} \\ \vdots \\ \beta_p^{(1)} \\ \beta_1^{(2)} \\ \vdots \\ \beta_p^{(2)} \end{bmatrix}}_{P \times 1}
+ \varepsilon
$$

N = # time points × # sub (300, 20)

P = # voxels × # subs (1000~30K, 20)

X is big, sparse.

$\boxed{\text{Solution available for sparse } X}$

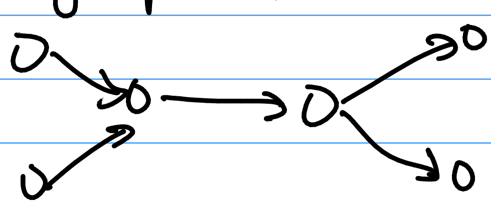
We can use group lasso to estimate graph edges. thus Graphical group Lasso (Friedman, 2010)

A few notes:

* Estimate sparseness also can estimate arc strengh (Gaël Varoquaux) only estimate sparseness.

arc strength $\Leftrightarrow$ β

* disregard temporal correlation (same with linear correlation)
* P independent regression problem $\Rightarrow$ potential to parallel
* choose λ (sparseness), cross validation.
* ~~sparcity should be smooth~~
* Use network as regressor for regression on clinical variable ?
* Negative correlation ?
* Can Lasso replace the regression on white matter and CSF ?
* Use flow-cut to analyze network ?
* How to generate synthetic data
generate graph first ?



$\Rightarrow$ generate sample ?

2. also estimate the sparcity of edges.

△ $\| \beta \|_K = (\beta' K \beta)^{\frac{1}{2}}$

△ Bayesian intepretation: regulizer as prior

Lasso ⇒ Laplacian prior.

$\boxed{\text{DTI into prior ?}}$

△ Post processing on the network

small-worldness ...

△ Multi-task Learning in ML community.