Group lasso with Overlap and Graph Lasso

Laurent Jacob[1,2]
Guillaume Obozinski[3,4]
Jean-Philippe Vert[1,2]

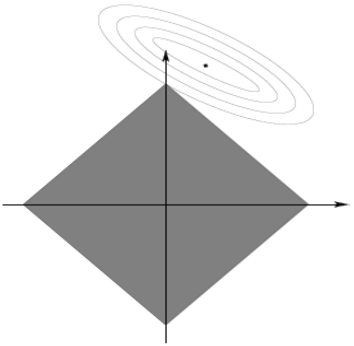
[1]Mines ParisTech, Centre for Computational Biology
[2]Institut Curie, INSERM U900
[3]Ecole Normale Supérieure
[4]INRIA – Willow project

16 juin 2009

Sparsity-inducing norms

Lasso

Well known that regularizing a learning problem by ℓ_1-norm induces sparse solutions (Tibshirani, 1996, Chen *et al.*, 1998).

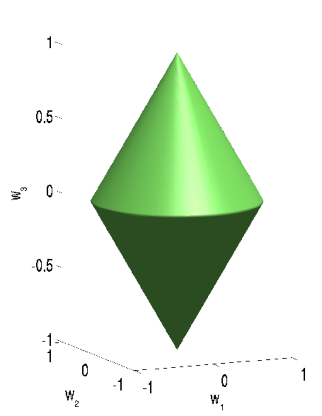


$$min_w L(w) + \lambda \|w\|_1.$$

Sparsity-inducing norms

Group lasso

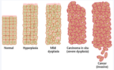
If groups of covariates are likely to be selected together, the ℓ_1/ℓ_2-norm induces sparse solutions *at the group level* (Yuan & Lin, 2006).



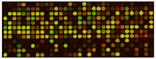
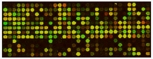
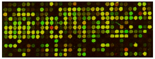
$$\min_w L(w) + \lambda \left(\|(w_1, w_2)\|_2 + \|w_3\|_2 \right).$$

Biological markers for cancer

Metastasis prognosis



Metastasizing tumors ?

Gene expression in tumor	Metastasis ?
	✓
⋮	⋮
	✗
	?

Predict metastasis, identify few predictive genes.

Biological markers for cancer

Gene selection

- X is the expression matrix of p genes for n tumors.

$$w = \begin{bmatrix} 0 \\ w_k \\ 0 \\ w_l \\ 0 \end{bmatrix}$$

- Learning with a ℓ_1-penalty favors a linear classifier $w \in \mathbb{R}^p$ involving *few genes*.
- Remark : may only select one of several correlated genes.
- After this selection, people often try to find enriched *functional groups*.

Biological markers for cancer

Overlapping groups

- We have prior information under the form of groups of genes with functional meaning (*e.g.* pathways).
- We would like to favor directly w involving few groups
  - Better interpretability.
  - Correlated genes typically in the same group, hence selected together.
  - Robustness to spurious gene selection.
- Group lasso originally proposed for disjoint groups.
- For overlapping groups, $\Omega_{group}(w) = \sum_{g \in \mathcal{G}} \|w_g\|_2$ is still a norm and has been considered for :
  - Hierarchical variable selection (Zhao *et al.* 2006, Bach 2008).
  - Structured sparsity (Jenatton *et al.* 2009).

Biological markers for cancer

Overlapping groups

- We have prior information under the form of groups of genes with functional meaning (*e.g.* pathways).
- We would like to favor directly w involving few groups
  - Better interpretability.
  - Correlated genes typically in the same group, hence selected together.
  - Robustness to spurious gene selection.
- Group lasso originally proposed for disjoint groups.
- For overlapping groups, $\Omega_{group}(w) = \sum_{g \in \mathcal{G}} \|w_g\|_2$ is still a norm and has been considered for :
  - Hierarchical variable selection (Zhao *et al.* 2006, Bach 2008).
  - Structured sparsity (Jenatton *et al.* 2009).

Biological markers for cancer

Overlapping groups

- We have prior information under the form of groups of genes with functional meaning (*e.g.* pathways).
- We would like to favor directly w involving few groups
  - Better interpretability.
  - Correlated genes typically in the same group, hence selected together.
  - Robustness to spurious gene selection.
- Group lasso originally proposed for disjoint groups.
- For overlapping groups, $\Omega_{group}(w) = \sum_{g \in \mathcal{G}} \|w_g\|_2$ is still a norm and has been considered for :
  - Hierarchical variable selection (Zhao *et al.* 2006, Bach 2008).
  - Structured sparsity (Jenatton *et al.* 2009).

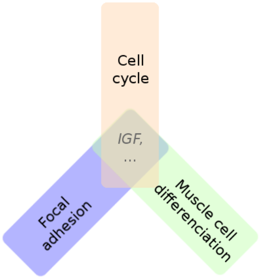
Biological markers for cancer

Overlapping groups

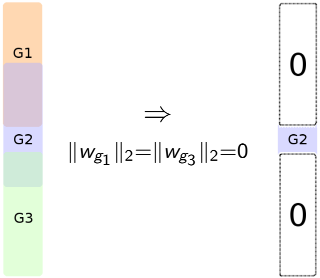
- We have prior information under the form of groups of genes with functional meaning (*e.g.* pathways).
- We would like to favor directly w involving few groups
  - Better interpretability.
  - Correlated genes typically in the same group, hence selected together.
  - Robustness to spurious gene selection.
- Group lasso originally proposed for disjoint groups.
- For overlapping groups, $\Omega_{group}(w) = \sum_{g \in \mathcal{G}} \|w_g\|_2$ is still a norm and has been considered for :
  - Hierarchical variable selection (Zhao *et al.* 2006, Bach 2008).
  - Structured sparsity (Jenatton *et al.* 2009).

Issue of using the group-lasso

- $\Omega_{group}(w) = \sum_g \|w_g\|_2$ sets groups to 0.
- One variable is selected $\Leftrightarrow$ all the groups to which it belongs are selected.



IGF selection $\Rightarrow$ selection of unwanted groups



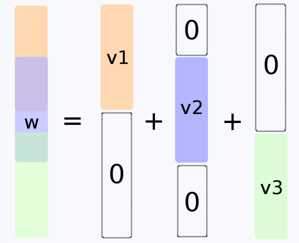
Removal of *any* group containing a gene $\Rightarrow$ the weight of the gene is 0.

Overlap norm

Overlap norm

Introduce latent variables v_g :

$$\begin{cases} \min_{w,v} L(w) + \lambda \sum_{g \in \mathcal{G}} \|v_g\|_2 \\ w = \sum_{g \in \mathcal{G}} v_g \\ \operatorname{supp}(v_g) \subseteq g. \end{cases}$$



Properties

- Resulting support is a *union* of groups in $\mathcal{G}$.
- Possible to select one variable without selecting all the groups containing it.
- Setting one v_g to 0 doesn't necessarily set to 0 all its variables in w.

Overlap norm

Overlap norm

$$\begin{cases} \min_{w,v} L(w) + \lambda \sum_{g \in \mathcal{G}} \|v_g\|_2 \\ w = \sum_{g \in \mathcal{G}} v_g \\ \operatorname{supp}(v_g) \subseteq g. \end{cases} = \min_w L(w) + \lambda \Omega_{overlap}(w)$$

with

$$\Omega_{overlap}(w) \triangleq \begin{cases} \min_{v} \sum_{g \in \mathcal{G}} \|v_g\|_2 \\ w = \sum_{g \in \mathcal{G}} v_g \\ \operatorname{supp}(v_g) \subseteq g. \end{cases} \qquad (*)$$

Property

- $\Omega_{overlap}(w)$ is a norm of w.
- $\Omega_{overlap}(.)$ associates to w a specific (not necessarily unique) decomposition $(v_g)_{g \in \mathcal{G}}$ which is the argmin of $(*)$.

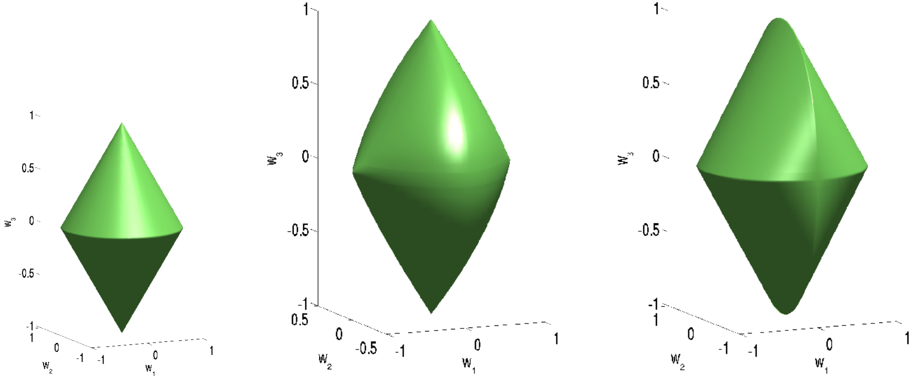
Equivalent formulation

Regular group-lasso in latent variable space

$$\begin{cases} \min_{w,v} L(Xw) + \lambda \sum_g \|v_g\|_2 \\ w = \sum_g v_g \\ \operatorname{supp}(v_g) \subseteq g. \end{cases} = \min_{\tilde v} L(\tilde X \tilde v) + \lambda \sum_g \|\tilde v_g\|_2$$

$$Xw = X.\begin{pmatrix}\tilde v_1\\ 0\end{pmatrix} + X.\begin{pmatrix}0\\ \tilde v_2\\ 0\end{pmatrix} + X.\begin{pmatrix}0\\ \tilde v_3\end{pmatrix} = (X_{g_1}, X_{g_2}, X_{g_3}).\begin{pmatrix}\tilde v_1\\ \tilde v_2\\ \tilde v_3\end{pmatrix} \triangleq \tilde X \tilde v.$$

Overlap and group unity balls



Balls for $\Omega^{\mathcal{G}}_{\text{group}}(\cdot)$ (middle) and $\Omega^{\mathcal{G}}_{\text{overlap}}(\cdot)$ (right) for the groups $\mathcal{G} = \{\{1,2\},\{2,3\}\}$ where w_2 is represented as the vertical coordinate. Left : group-lasso ($\mathcal{G} = \{\{1,2\},\{3\}\}$), for comparison.

Overlap norm

Consistency in group support

- Let $\bar{w}$ be the true parameter vector.
- Assume that there exists a unique decomposition $\bar{v}_g$ such that $\bar{w} = \sum_g \bar{v}_g$ and $\Omega_{\text{overlap}}^{\mathcal{G}}(\bar{w}) = \sum \|\bar{v}_g\|_2$.
- Consider the regularized empirical risk minimization problem $L(w) + \lambda \Omega_{\text{overlap}}^{\mathcal{G}}(w)$.

Then

- under appropriate mutual incoherence conditions on X,
- as $n \to \infty$,
- with very high probability,

the optimal solution $\hat{w}$ admits a unique decomposition $(\hat{v}_g)_{g \in \mathcal{G}}$ such that

$$\{g \in \mathcal{G} | \hat{v}_g \neq 0\} = \{g \in \mathcal{G} | \bar{v}_g \neq 0\}.$$

Overlap norm

Consistency in group support

- Let $\bar{w}$ be the true parameter vector.
- Assume that there exists a unique decomposition $\bar{v}_g$ such that $\bar{w} = \sum_g \bar{v}_g$ and $\Omega^{\mathcal{G}}_{\text{overlap}}(\bar{w}) = \sum \|\bar{v}_g\|_2$.
- Consider the regularized empirical risk minimization problem $L(w) + \lambda \Omega^{\mathcal{G}}_{\text{overlap}}(w)$.

Then

- under appropriate mutual incoherence conditions on X,
- as $n \to \infty$,
- with very high probability,

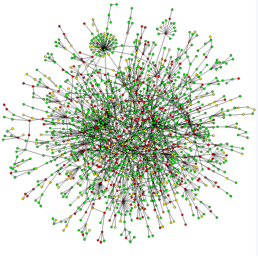
the optimal solution $\hat{w}$ admits a unique decomposition $(\hat{v}_g)_{g \in \mathcal{G}}$ such that

$$\{g \in \mathcal{G} | \hat{v}_g \neq 0\} = \{g \in \mathcal{G} | \bar{v}_g \neq 0\} .$$

.

Graph lasso

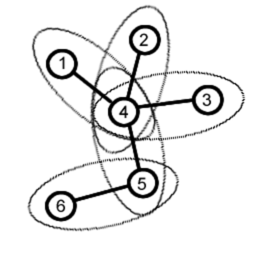
- Other types of biological priors can be represented as graphs (protein interaction, gene regulation...).



- In that case, it is reasonable to expect that relevant genes form connected components in such a graph.
- Moreover, these components might be used as groups of potential drug targets and uncover biological processes relevant for metastasis.

Graph lasso

- Consider groups that are subgraphs whose union would give such connected components (*e.g.*, edges E).

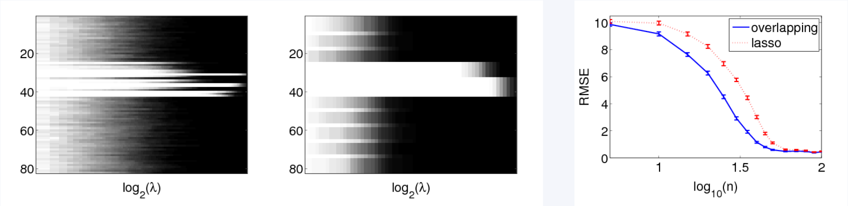


- $\Omega_{\text{graph}}(w) = \min_{v \in \mathcal{V}_E} \sum_{e \in E} \|v_e\| \quad \text{s.t.} \sum_{e \in E} v_e = w, \ \text{supp}\,(v_e) = e\,.$

Results

Synthetic data : overlapping groups

- 10 groups of 10 variables with 2 variables of overlap between two successive groups :$\{1, \dots, 10\}, \{9, \dots, 18\}, \dots, \{73, \dots, 82\}$.
- Support : union of $4th$ and $5th$ groups.
- Learn from 100 training points.



Frequency of selection of each variable with the lasso (left) and $\Omega^{\mathcal{G}}_{\text{overlap}}(.)$ (middle), comparison of the RMSE of both methods (right).

Results

Breast cancer data

- Gene expression data for $8,141$ genes in 295 breast cancer tumors.
- Canonical pathways from MSigDB containing 639 groups of genes, 637 of which involve genes from our study.

Method	ℓ_1	$\Omega^{\mathcal{G}}_{\text{overlap}}(.)$
Error	0.38 ± 0.04	0.36 ± 0.03
$\sharp$ path.	$148, 58, 183$	$6, 5, 78$
Prop. path.	$0.32, 0.14, 0.41$	$0.01, 0.01, 0.17$

- Graph on the genes.

Method	ℓ_1	$\Omega_{graph}(.)$
Error	0.39 ± 0.04	0.36 ± 0.01
Av. size c.c.	$1.1, 1, 1.0$	$1.3, 1.4, 1.2$

Conclusion

Summary

- Generalization of the group-lasso penalty leading to sparsity patterns which are *unions* of overlapping groups.
- Helps to recover sparse connected patterns in a graph.
- Group-consistency conditions.
- Encouraging results on breast cancer data.

Future works

- Comparison with Ω_{group} when both retrieve the same class of patterns (*e.g.* graphs).
- Weighted penalty (group sizes, overlap sizes).
- More general consistency conditions.

Overlap norm

Dual formulation

1.

$$\Omega_{overlap}(w) = \begin{cases} \inf_v \sum_g \|v_g\|_2 \\ w = \sum_g v_g \\ \operatorname{supp}(v_g) \subseteq g. \end{cases} = \begin{cases} \sup_\alpha \alpha^\top w \\ \forall g, \|\alpha_g\|_2 \leq 1 \end{cases} \quad (1)$$

2. A vector $\alpha \in \mathbb{R}^p$ is a solution of (1) if and only if there exists $\mathbf{v} = (v_g)_{g \in \mathcal{G}} \in \mathbf{V}(w)$ such that :

$$\forall g \in \mathcal{G}\,, \quad \text{if } v_g \neq 0, \; \alpha_g = \frac{v_g}{\| v_g \|} \; \text{ else } \| \alpha_g \| \leq 1 \quad (2)$$

3. Conversely, a $\mathcal{G}$-tuple of vectors $\mathbf{v} = (v_g)_{g \in \mathcal{G}} \in \mathcal{V}_\mathcal{G}$ such that $w = \sum_g v_g$ is a solution to (1) if and only if there exists a vector $\alpha \in \mathbb{R}^p$ such that (2) holds.

Overlap norm

Consistency

If we assume that

1. (H1) $\Sigma := \frac{1}{n} X^\top X \succ 0$
2. (H2) There exists a neighborhood of $\bar{w}$ in which the decomposition in v is unique,

then

$$\forall g \in \mathcal{G}_2, \; \|\Sigma_{gJ_1} \Sigma_{J_1 J_1}^{-1} \alpha_{J_1}(\bar{w})\| \leq 1 \tag{C1}$$

$$\forall g \in \mathcal{G}_2, \; \|\Sigma_{gJ_1} \Sigma_{J_1 J_1}^{-1} \alpha_{J_1}(\bar{w})\| < 1 \tag{C2}$$

are respectively necessary and sufficient for the minimization of

$$\min_{w \in \mathbb{R}^p} R(w) + \lambda \Omega^{\mathcal{G}}_{\text{overlap}}(w) \;, \tag{3}$$

to estimate consistently the group-support of $\bar{w}$.

Overlap norm

Consistency : Remark

- Consistency conditions for $\Omega^{\mathcal{G}}_{\text{overlap}}(.)$:

$$\forall g \in \mathcal{G}_2,\ \|\Sigma_{gJ_1}\Sigma^{-1}_{J_1J_1}\alpha_{J_1}(\bar{w})\| \leq 1 \qquad \text{(C1)}$$

$$\forall g \in \mathcal{G}_2,\ \|\Sigma_{gJ_1}\Sigma^{-1}_{J_1J_1}\alpha_{J_1}(\bar{w})\| < 1 \qquad \text{(C2)}$$

- Consistency conditions for group-lasso (Bach *et al.*, 2008) :

$$\forall g \in \mathcal{G}_2,\ \|\Sigma_{gJ_1}\Sigma^{-1}_{J_1J_1}\mathrm{Diag}(1/\|\bar{w}_{J_1,i}\|_2)_i\bar{w}_{J_1}\| \leq 1 \qquad \text{(C1)}$$

$$\forall g \in \mathcal{G}_2,\ \|\Sigma_{gJ_1}\Sigma^{-1}_{J_1J_1}\mathrm{Diag}(1/\|\bar{w}_{J_1,i}\|_2)_i\bar{w}_{J_1}\| < 1 \qquad \text{(C2)}$$

- No closed form for $\alpha(\bar{w})$ in the general case.
- If there is no overlap, we recover the group-lasso result.