Co-data course: GRridge

Mark van de Wiel[1,2,3], Putri Novianti[1], Magnus Münch[1,4]

[1]Dep of Epidemiology and Biostatistics, VU University medical center

[2]Dep of Mathematics, VU University, Amsterdam, NL

[3]MRC Biostatistics unit, Cambridge University, UK (Visiting Fellow)

[4]Mathematical Institute, Leiden University, NL

Our group: www.bigstatistics.nl

Setting

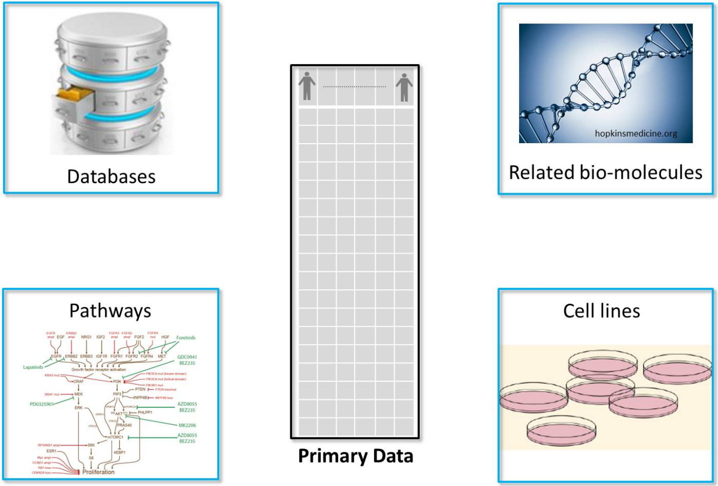
- **Prediction or Diagnosis**
- **Primary data**
  - Variables $i = 1, \ldots, p$; Individuals $j = 1, \ldots, n$; $p > n$
  - Focus on binary response Y_j (e.g. case vs control)
  - Measurements $\mathbf{X}_j = (X_{1j}, \ldots, X_{pj})$
  - Goal: find f such that $Y_j \approx f(\mathbf{X}_j)$
  - Here, f: *logistic regression*
  - Some form of regularization required
- **Focus**
  - Differential regularization based on prior information: **co-data**

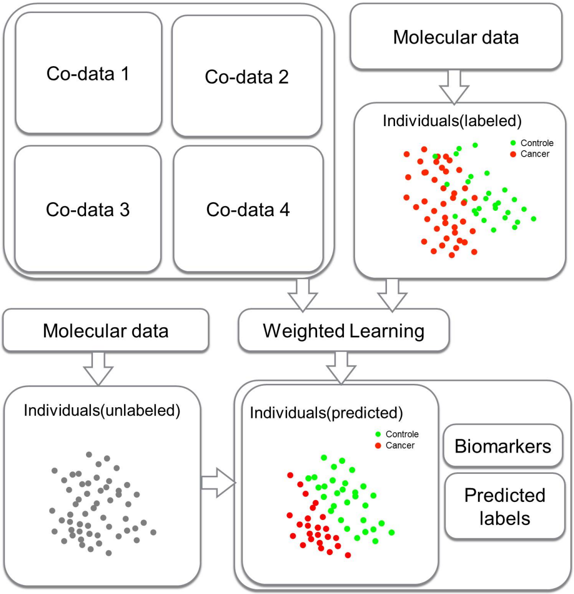
Co-data

Definition Co-data: any information on the *variables* that does not use the response labels of the primary data

Co-data

Definition Co-data: any information on the *variables* that does not use the response labels of the primary data





Co-data 1
Co-data 2
Co-data 3
Co-data 4
Molecular data
Individuals(labeled)
Controle
Cancer
Molecular data
Weighted Learning
Individuals(unlabeled)
Individuals(predicted)
Controle
Cancer
Biomarkers
Predicted labels

Use of co-data

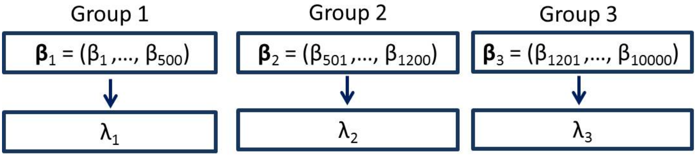
Groups: Co-data determine G prior groups of variables

Idea: Use different penalty weights $\boldsymbol{\lambda} = \lambda_1, \ldots, \lambda_G$ across G co-data-based groups.

Use of co-data

Groups: Co-data determine G prior groups of variables

Idea: Use different penalty weights $\boldsymbol{\lambda} = \lambda_1, \ldots, \lambda_G$ across G co-data-based groups. $G = 3$:



E.g. Ridge: $\text{argmax}_{\beta}\{\mathcal{L}(\mathbf{Y};\boldsymbol{\beta}) - \sum_{g=1}^{G} \lambda_g ||\boldsymbol{\beta}_g||_2\}$

$\rightarrow$ **CV** not attractive
$\rightarrow$ `GRridge` estimates $\boldsymbol{\lambda}$ by Empirical Bayes (EB)

Fitting

Likelihood contains term $\mathbf{X}'\beta$. Write $\lambda_g = \lambda_g'\lambda$

$$\text{argmax}_{\boldsymbol{\beta}}\{\mathcal{L}(\mathbf{Y};\mathbf{X}'\boldsymbol{\beta}) - \lambda\sum_{g=1}^{G}\lambda_g'||\boldsymbol{\beta}_g||_2\}$$
$$\equiv$$
$$\text{argmax}_{\tilde{\boldsymbol{\beta}}}\{\mathcal{L}(\mathbf{Y};\tilde{\mathbf{X}}'\tilde{\boldsymbol{\beta}}) - \lambda||\tilde{\boldsymbol{\beta}}||_2\},$$

where

$$\tilde{\mathbf{X}} = \mathbf{X}(\text{diag}(\boldsymbol{\lambda}'))^{-1/2} \text{ and } \tilde{\boldsymbol{\beta}}_g = \boldsymbol{\beta}_g(\lambda_g')^{1/2}$$

$\rightarrow$ Existing software used for fitting (`glmnet`, `penalized`)

Iteration

One may iterate the hyper-parameter estimation:

$$\text{argmax}_{\tilde{\boldsymbol{\beta}}}\{\mathcal{L}(\mathbf{Y};\tilde{\mathbf{X}}'\tilde{\boldsymbol{\beta}}) - \lambda||\tilde{\boldsymbol{\beta}}||_2\},\ \tilde{\mathbf{X}} = \mathbf{X}(\text{diag}(\boldsymbol{\lambda}'))^{-1/2}$$

$\rightarrow$ Same group-structure on $\tilde{\boldsymbol{\beta}}$:

$$\text{argmax}_{\boldsymbol{\beta}}\{\mathcal{L}(\mathbf{Y};\tilde{\mathbf{X}}'\tilde{\boldsymbol{\beta}}) - \lambda\sum_{g=1}^{G}\lambda''_g||\tilde{\boldsymbol{\beta}}_g||_2\}$$

Effective penalty multiplier: $\lambda'_g\lambda''_g$

Convergence is monitored by cross-validated likelihood

Multiple partitions I

Current solution: Iterative. Suppose first partition $\mathcal{G}$ has been used.

Different group-structure $\mathcal{H}$ on $\tilde{\boldsymbol{\beta}}$:

$$\text{argmax}_{\boldsymbol{\beta}}\{\mathcal{L}(\mathbf{Y};\tilde{\mathbf{X}}'\tilde{\boldsymbol{\beta}}) - \lambda \sum_{h=1}^{H} \lambda''_h ||\tilde{\boldsymbol{\beta}}_h||_2\}$$

Effective penalty multiplier for variable k: $\lambda'_{g(k)}\lambda''_{h(k)}$

Disadvantage: order may matter, although partitions are interwoven with iterations

Multiple partitions II

A nicer solution is implemented in `gren`:

$$\hat{\boldsymbol{\beta}} := \underset{\boldsymbol{\beta}}{\text{argmax}} \ell(\mathbf{Y}; \boldsymbol{\beta}) - \frac{\lambda_1}{2} \sum_{g=1}^{G} \sum_{h=1}^{H} \sqrt{\lambda'_g \lambda''_h} \sum_{\substack{j \in \mathcal{G}_g \\ \cap \mathcal{H}_h}} |\beta_j| - \frac{\lambda_2}{2} \sum_{g=1}^{G} \sum_{h=1}^{H} \lambda'_g \lambda''_h \sum_{\substack{j \in \mathcal{G}_g \\ \cap \mathcal{H}_h}} \beta_j^2$$

Overlapping groups*

Prior modeled as:

$$\beta_k \sim N(0, \sum_{g_k \in \mathcal{J}_k} \sigma^2_{g_k}/|\mathcal{J}_k|)$$

$\rightarrow$ Set of G estimation equations $\rightarrow \hat{\sigma}^2_\ell$

Set $\lambda_\ell \propto \hat{\sigma}_\ell^{-2}$

*See Novianti et al. (2017), *Bioinformatics*

Data

Data set 1

- Methylation data on cervical tissue, $p = 40,000$, $n = 37$
- Classification: Normal vs CIN3 (Precursor)
- Co-data: Location of the probes, 6 groups
- Reference: Farkas et al. (2013), *Epigenetics*
- Also analysed in: Van de Wiel et al. (2016), *Stat Med*

Data set 2

- RNAseq data from *blood* platelets, $p = 18,410$, $n = 81$
- Classification: Breast vs. Colon cancer
- Co-data: Several (see course text)
- Reference: Best et al. (2015), *Cancer Cell*
- Also analysed in: Novianti et al. (2017), *Bioinformatics*

Getting started

Instructions and Exercises:
See `https://magnusmunch.github.io/co-data_learning/`