Modeling high-dimensional interaction problems with the pliable lasso

Asenso Theophilus Quachie (t.q.asenso@medisin.uio.no)

Oslo Centre for Biostatistics and Epidemiology, University of Oslo

Joint work with Manuela Zucknick

Motivation

- Modelling **interactions in high-dimensional** data is a notoriously difficult problem.
- Analyzing high-dimensional data with conventional tools is very challenging.
  - Most existing models cannot easily handle **high-dimensional data and many interactions**.
  - Most methods often make **strong assumptions**, e.g. strong hierarchy.
- **Pliable lasso** [Tibshirani and Friedman, 2020] relaxes the strong hierarchy assumption
- We present two extensions of pliable lasso, including **the MADMM algorithm for multi-response data**.
- We apply MADMM to cancer **drug sensitivity screening** data for personalised cancer therapy .

Contents

Introduction to the pliable lasso model

$y \in \mathbb{R}^N$, $X \in \mathbb{R}^{N \times p}$ and $Z \in \mathbb{R}^{N \times K}$.The pliable lasso [Tibshirani and Friedman, 2020] model is given as;

$$\begin{aligned} \hat{y} &= \beta_0 \mathbf{1} + Z\theta_0 + \sum_{j=1}^{p} X_j(\beta_j \mathbf{1} + Z\theta_j) \\ &= \beta_0 + Z\theta_0 + X\beta + \sum_{j=1}^{p} (X_j \odot Z)\theta_j, \end{aligned} \tag{1}$$

where $(X_j \odot Z)$ denoting the $N \times K$ matrix formed by multiplying each column of Z component-wise by the column vector X_j.

Introduction to the pliable lasso model

The pliable lasso objective function

$$M(\beta_0, \theta_0, \beta, \theta) = \frac{1}{2N} \sum_i (y_i - \hat{y}_i)^2 + (1-\alpha)\lambda \sum_{j=1}^{p} \overbrace{(\|(\beta_j, \theta_j)\|_2 + \|\theta_j\|_2)}^{\text{Overlapping group}} + \alpha\lambda \sum_{j,k} |\theta_{j,k}| \quad (2)$$

- y_i is the element of the fitted model $\beta_0 \mathbf{1} + Z\theta_0 + \sum_{j=1}^{p} X_j(\beta_j \mathbf{1} + Z\theta_j)$.
- Overlapping group ensures **(asymmetric) weak hierarchy constraint**.

A Bayesian representation the Pliable lasso model

A Bayesian Pliable Lasso

$$y|\beta_0,\beta,\theta_0,\theta,\sigma^2 \sim \mathcal{N}(\beta_0\mathbf{1} + Z\theta_0 + X\beta + \sum_{j=1}^{p} X_j \circ Z\theta_j, \sigma^2\mathbb{I}_N), \tag{3}$$

$$\beta_0 \sim \mathcal{N}(0, c^2), \quad \theta_0|\tau^2_{\theta_0},\sigma^2 \sim \mathcal{N}(0, \tau^2_{\theta_0}\sigma^2\mathbb{I}_K), \quad \sigma^2 \sim \text{Inverse Gamma}(\lambda_1,\lambda_2), \quad \sigma^2 > 0,$$

$$\beta_j|\eta_j,\tau^2_{\beta_j},\sigma^2 \sim \eta_j\mathcal{N}(0,\ \tau^2_{\beta_j}\sigma^2) + (1-\eta_j)\delta_0(\beta_j),$$

$$\theta_{jk}|\eta_j = 1,\gamma_{jk},\tau^2_{\theta_{jk}},\sigma^2 \sim \gamma_{jk}\mathcal{N}(0,\ \tau^2_{\theta_{jk}}\sigma^2) + (1-\gamma_{jk})\delta_0(\theta_{jk}),$$

where c is a nonzero constant, $\tau^2_{\theta_0} \sim \text{Gamma}((K+1)/2, v^2/2)$, $\tau^2_{\beta_j} \sim \text{Gamma}(1,(1-\alpha)^2\lambda^2/2)$, $\alpha \in [0,1]$ and $\delta_0(\cdot)$ is the Dirac delta function, $\tau_{\theta_{jk}} \sim \text{Gamma}(1,\alpha^2\lambda^2/2)$.
NB conditioning θ_{jk} only on $\eta_j = 1$ guarantees asymmetric weak hierarchical interactions.

$$\eta_j \sim \text{Bernoulli}(\rho), \gamma_{jk} \sim \text{Bernoulli}(\zeta),$$
$$\rho \sim \text{Beta}(a_\rho, b_\rho), \zeta \sim \text{Beta}(a_\zeta, b_\zeta).$$

A Bayesian Pliable Lasso

Table: Results from the three examples, based on 50 replications.

Model	$(1/(p+p\times K))\|\hat{\mathbf{B}}-\mathbf{B}\|_1$	Sensitivity	Specificity	Accuracy	FPR	# Non-zeros	Test error
Scenario 1	$p=20$, $SNR=2$						
BpLasso	0.021 (0.010)	0.993 (0.047)	0.997 (0.007)	0.996 (0.007)	0.003 (0.007)	4 (2)	1.663 (2.824)
pLasso	0.059 (0.023)	1.000 (0.000)	0.743 (0.133)	0.758 (0.125)	0.257 (0.133)	12 (17)	3.127 (1.142)
Scenario 2	$p=20$, $SNR=2$						
BpLasso	0.021 (0.011)	0.980 (0.054)	0.998 (0.005)	0.997 (0.005)	0.002 (0.005)	4 (2)	0.773 (0.492)
pLasso	0.097 (0.026)	1.000 (0.000)	0.659 (0.137)	0.680 (0.128)	0.341 (0.137)	16 (22)	1.839 (0.611)
Scenario 3	$p=20$, $SNR=6$						
BpLasso	0.018 (0.005)	1.000 (0.008)	0.970 (0.024)	0.975 (0.021)	0.030 (0.024)	11 (8)	0.265 (0.119)
pLasso	0.034 (0.007)	1.000 (0.000)	0.530 (0.140)	0.615 (0.115)	0.469 (0.140)	17 (39)	0.459 (0.147)
Scenario 3	$p=500$, $SNR=6$						
BpLasso	0.001 (0.000)	0.970 (0.068)	1.000 (0.000)	0.999 (0.001)	0.000 (0.000)	10 (7)	0.313 (0.312)
pLasso	0.004 (0.001)	0.941 (0.087)	0.962 (0.018)	0.962 (0.018)	0.038 (0.018)	91 (19)	2.277 (0.534)
Scenario 3	$p=20$, $SNR=2$						
BpLasso	0.034 (0.010)	0.910 (0.082)	0.959 (0.030)	0.950 (0.029)	0.041 (0.029)	12 (7)	1.047 (0.466)
pLasso	0.050 (0.011)	0.993 (0.021)	0.566 (0.130)	0.643 (0.106)	0.433 (0.129)	17 (36)	1.129 (0.325)

Multi-response problem with main and interaction effects

Drug dose response

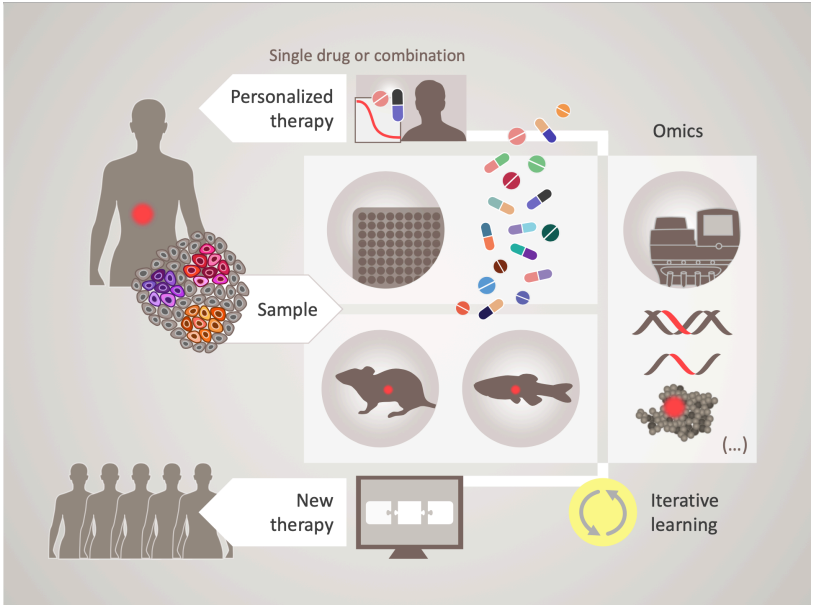
$$\text{N cell lines} \overbrace{\begin{bmatrix} | & & | \\ y_{\cdot 1} & \dots & y_{\cdot D} \\ | & & | \end{bmatrix}}^{\text{drug sensitivity}} = Y$$

Genetic features

$$\text{N cell lines} \overbrace{\begin{bmatrix} | & & | \\ X_{\cdot 1} & \dots & X_{\cdot p} \\ | & & | \end{bmatrix}}^{\text{gene expression}} = X$$

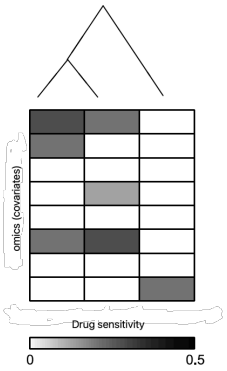
Interactions

$$\text{N cell lines} \overbrace{\begin{bmatrix} | & & | \\ Z_{\cdot 1} & \dots & Z_{\cdot K} \\ | & & | \end{bmatrix}}^{\text{Cancer type}} = Z$$



slide by Kjetil Taskén

Possible structures to study



- Structures in the response matrix ([Kim and Xing, 2012], [Li et al., 2015]) for example correlations between drug responses due to similar chemical properties, drug target, drug functions, etc
- Structures within the covariates or with a set of modifying variables ([Li et al., 2015], [Tibshirani and Friedman, 2020]) for example gene-to-gene interactions, gene-to-cancer type interactions, correlated genes, etc

Model specification

- Let $B \in \mathbb{R}^{D \times p \times (K+1)}$.
- The j^{th} row of B_d defined as $B_{jd} = [\boldsymbol{\beta}_{jd}, \boldsymbol{\theta}_{jd}] \in \mathbb{R}^{K+1}$.
- Let W be an $N \times p \times (1 + K)$

$$W_{i,j,k} = \begin{cases} X_{ij}Z_{ik} & \text{for} \quad k \neq 1 \\ X_{ij} & \text{for} \quad k = 1, \end{cases} \tag{4}$$

$k = 1, 2, \ldots, K + 1.$

$$\hat{Y} = \mathbf{1}\beta_0^T + Z\boldsymbol{\theta} + W * B, \tag{5}$$

where $W * B = [W * B_1 : W * B_2 : \ldots : W * B_D]$ to denote $N \times D$ matrix whose i, d element takes the form

$$(W * B)_{id} = \sum_{j=1}^{p} \sum_{k=1}^{K+1} W_{i,j,k} B_{jkd}, \quad i = 1, 2, \ldots N, \quad d = 1, 2, \ldots, D. \tag{6}$$

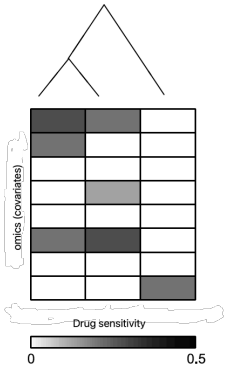
Model specification (with plasso)

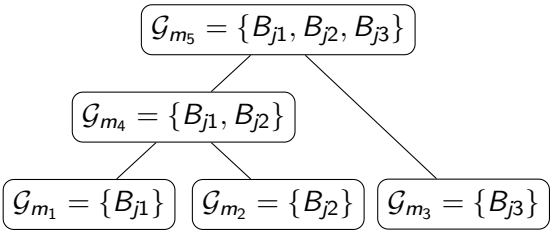
- $B \in \mathbb{R}^{D \times p \times (K+1)}$.

The general multi-response pliable lasso model can be written as

$$\min_{B \in \mathbb{R}^{D \times p \times (1+K)}} \quad \frac{1}{2N} \| Y - \hat{Y} \|_F^2 + \sum_{d=1}^{D} \left[(1-\alpha)\lambda \sum_{j=1}^{p} (\| B_{jd} \|_2 + \| B_{j(-1)d} \|_2) + \alpha\lambda \sum_{j=1}^{p} \| B_{j(-1)d} \|_1 \right] \quad (7)$$

Model specification (with tree lasso)





The simplified version of [Kim and Xing, 2012] is;

$$\min_{B \in \mathbb{R}^{D \times p \times (1+K)}} \frac{1}{2N} \| Y - \hat{Y} \|_F^2 + \lambda \sum_{j=1}^{p} \sum_{m \in M} w_m \| B_j^{\mathcal{G}_m} \|_2 + \lambda \sum_{j=1}^{p} \sum_{d=1}^{D} w_d |B_{jd}|. \quad (8)$$

Our approach

Combining (7) and (8);

$$\min_{B\in\mathbb{R}^{D\times p\times(1+K)}} \frac{1}{2N}\|Y-\hat{Y}\|_F^2+\lambda_1\sum_{j=1}^{p}\sum_{m\in M} w_m\|B_j^{\mathcal{G}_m}\|_2 + \sum_{d}^{D}\left[(1-\alpha)\lambda_2\sum_{j=1}^{p}(\|B_{jd}\|_2+\|B_{j(-1)d}\|_2)+\alpha\lambda_2\sum_{j=1}^{p}\|B_{j(-1)d}\|_1\right]. \quad (9)$$

- We use **ADMM**: "The **alternating direction method of multipliers (ADMM)** is an algorithm that solves convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle. It has recently found wide application in a number of areas." (https://stanford.edu/ boyd/admm.html)

Our approach using ADMM [Boyd et al., 2011]

- Introduce three auxiliary variables $\mathbf{E}$, $\mathbf{V}$, $\mathbf{Q}$
- Let $\mathbf{E} \in \mathbb{R}^{(\sum_{m\in M}|\mathcal{G}_m|)\times(p+p\times K)}$;
- $\mathbf{E}_j \in \mathbb{R}^{\sum_{m\in M}|\mathcal{G}_m|} \times (K+1)$ to represent the groups formed by internal nodes M
- $\mathbf{E}^{\mathcal{G}_m} \in \mathbb{R}^{|\mathcal{G}_m|\times(p+p\times K)}$ being the block representation of group $\mathcal{G}_m$.
- $\mathcal{B}_j = [B_j^{\mathcal{G}_1 T}, B_j^{\mathcal{G}_2 T}, \ldots, B_j^{\mathcal{G}_M T}]^T \in \mathbb{R}^{\sum_{m\in M}|\mathcal{G}_m|\times(1+K)}$.
- Let $\{ B_{jd}^{g_s} \in \mathbb{R}^{k_s}; s = 1, 2\}$ be the grouping of elements of B_{jd}, where g_s with $\bigcup_s g_s \subseteq \{1, 2, \ldots, K\}$ is an index set corresponding to the s^{th} group
- $B_{jd}^{g_s}$ denotes the subvector of B_{jd}, indexed by g_s and k_s represents the size of g_s.
- $\mathbf{V}$ is an array of size D, $\mathbf{V}_d \in \mathbb{R}^{p\times\tilde{K}}$, with each j^{th} row vector $\mathbf{V}_{jd} \in \mathbb{R}^{\tilde{K}}$
- $\tilde{K} = \sum_{s=1}^{2} k_s \geq K$ and $\mathbf{V}_{jj}^1 = B_{jj}^{g_1}$ and $\mathbf{V}_{jd}^2 = B_{jd}^{g_2}$.
- Let $\mathbf{Q} \in \mathbb{R}^{D\times(p\times K)}$) be an array.

Our approach using ADMM [Boyd et al., 2011]

$$\mathcal{L}(B, \mathbf{E}, \mathbf{V}, \mathbf{Q}, \mathbf{H}, \mathbf{O}, \mathbf{P}) = \frac{1}{2N}\|Y - \hat{Y}\|_F^2 + \lambda_1 \sum_{j=1}^{p} \sum_{m \in M} w_m \|\mathbf{E}_j^{\mathcal{G}_m}\|_2$$

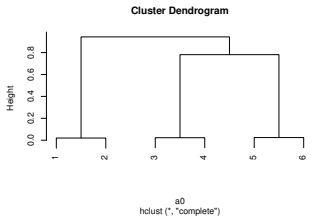
$$+ \sum_d (1-\alpha)\lambda_2 \sum_{j=1}^{p} \sum_s \|\mathbf{V}_{jd}^s\|_2 + \alpha\lambda_2 \sum_{j=1}^{p} \|\mathbf{Q}_{jd}\|_1 + \sum_j \mathbf{H}_j (\tilde{\tilde{B}}_j - \mathbf{E}_j)^T$$

$$+ \sum_d \sum_j \mathbf{O}_{jd} (\tilde{B}_{jd} - \mathbf{V}_{jd})^T + \sum_d \langle \mathbf{P}_d, B_d - \mathbf{Q}_d \rangle$$

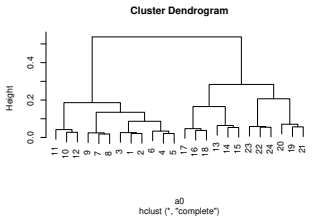
$$+ \frac{\rho}{2} \sum_j \|\tilde{\tilde{B}}_j - \mathbf{E}_j\|_2^2 + \frac{\rho}{2} \sum_d \sum_j \|\tilde{B}_{jd} - \mathbf{V}_{jd}\|_2^2 + \frac{\rho}{2} \sum_d \|B_d - \mathbf{Q}_d\|_2^2. \quad (10)$$

Simulation studies

$D = 7, p = 500, K = 4, N = 100$



$D = 24, p = 150, 500, K = 4, N = 100$



Simulated correlation structure of D drug response variables across N cell lines for simulated data set 1 (left) and 2 (right)."

Results for simulated data set 1

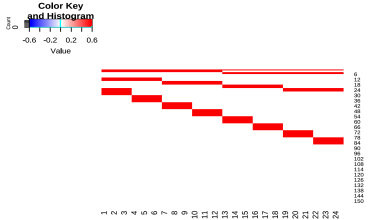
Table: Results from the simulated data 1 **without strong hierarchical structure** in the response

Model	$(1/Dp)\|\|\hat{\beta} - \beta\|\|_1$	Sensitivity	Specificity	Non-zero coefficient	Test error
Plasso	0.0027	1	0.940	208	19.835
Tree lasso	0.011	1	0.668	1016	30.927
MADMMplasso	0.0024	1	0.962	151	21.997

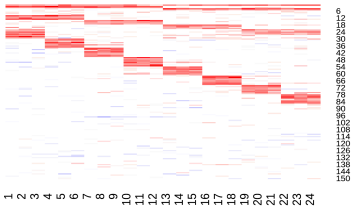
[1] Non zero Coefficient is the non zero main effects out of $p \times D = 500 \times 7 = 3500$.

[2] Test error is MSE in independent test data set.

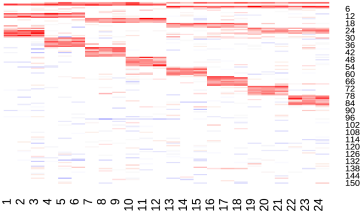
Results for simulated data set 2 (coefficient matrix)



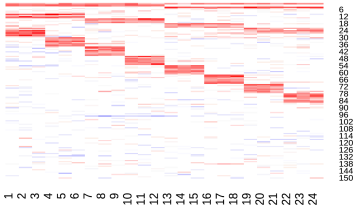
a True structure



b MADMMplasso



c plasso



d Tree lasso

Results for simulated data set 2

Table: Results from the simulated data 2 with **strong hierarchical structure** in the response

Model	$(1/Dp)\|\|\hat{\beta} - \beta\|\|_1$	Sensitivity	Specificity	Non-zero coefficient	Test error
$p = 150$					
Plasso	0.033	0.986	0.802	998	2.619
Tree lasso	0.037	1	0.747	987	2.461
MADMMplasso	0.030	1	0.794	1028	2.215
$p = 500$					
Plasso	0.012	0.892	0.922	1234	4.865
Tree lasso	0.02	0.994	0.795	2744	2.869
MADMMplasso	0.010	0.992	0.923	1276	2.202

[1] Non zero Coefficient is the non zero main effects out of $p \times D = 150 \times 24 = 3600$ or $500 \times 24 = 12000$.

[2] Test error is MSE in independent test data set.

Real data

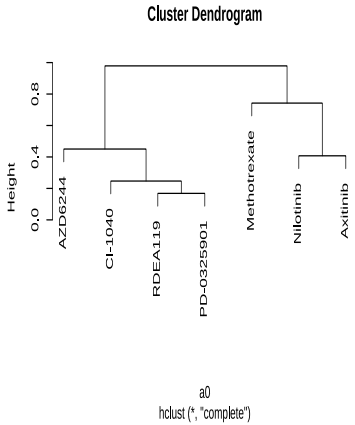
'Genomics of drug sensitivity in cancer' [Garnett et al., 2012]

- Large-scale pharmacogenomic study with $N = 498$ cell lines and $D = 97$ drugs (we used 7 drugs).
- Outcome data: $\log(IC_{50})$ from dose-response experiments
- Random draws of 80% cell lines as training data and 20% as validation data.
- Input data: Z as cancer types (13 cancer types, $K = 12$), X as mRNA expression (p=2602)

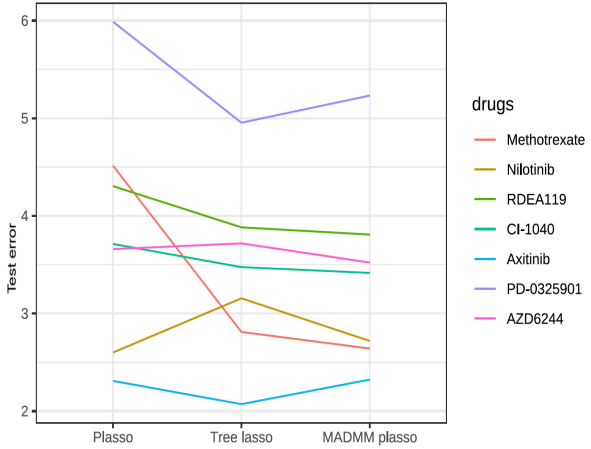
Drug information

- **PD-0325901, RDEA119, CI-1040, AZD6244**: MEK1 inhibitors with highly correlated IC50 values.
- **Methotrexate:** general cytotoxic drug not targeted to specific genes/pathways
- **Nilotinib:** inhibits the BCR-ABL fusion gene characteristic for chronic myeloid leukemia. Related to Axitinib (smaller effect)

Real data



e Correlation structure of 7 drug response variables across 400 cell lines



f Test error

Real data

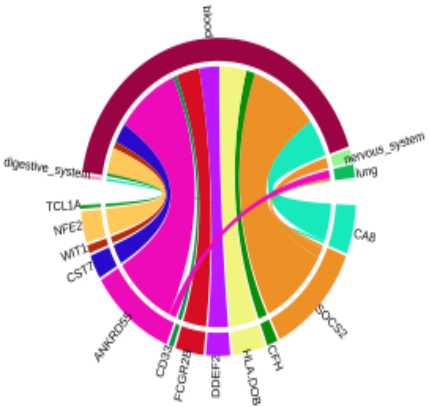
GDSC [Garnett et al., 2012]

Table: Results from the GDSC data

Model	Non zero coefficient	Test error
$p = 2602$		
Plasso	351	3.869
Tree lasso	603	3.438
MADMMplasso	928	3.380

[1] Non zero Coefficient is the non zero main effects out of $p \times D = 2602 \times 7 = 18214$.
[2] Test error is MSE in independent test data set.

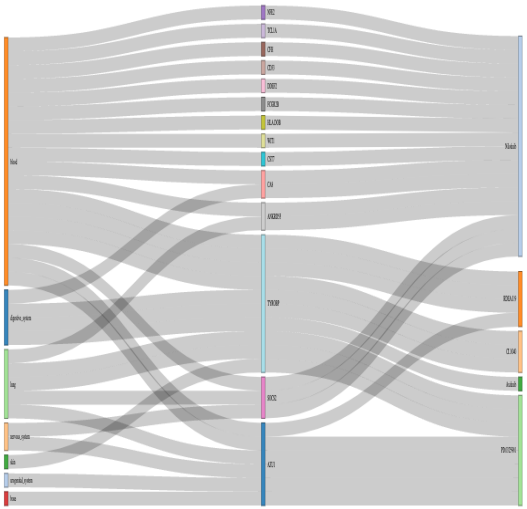
Real data results: Selected interaction effects for Nilotinib



Suppressor of cytokine signaling 2 (SOCS2) is involved in the signal transduction cascades in CML cells [Schultheis et al., 2002]

Real data results: Summary of all selected interaction effects

GDSC [Garnett et al., 2012]



Summary

- We have considered problems involving main and interaction effects.
- The procedure involved the implementation of the **pliable lasso penalty**.
- Our extensions
  - **A Bayesian pliable lasso** with **with potential to handle false positives**
  - **Multi-response problem** with **tree-guided structure**.
  - The implementation of the **ADMM algorithm** made it possible to handle the overlapping groups in both the covariates and the responses.
  - The R package (**MADMMplasso**) is publicly available on https://github.com/ocbe-uio/MADMMplasso

Email: t.q.asenso@medisin.uio.no

This work received funding from the European Union's Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie Actions Grant, agreement No. 80113 (Scientia fellowship). It has also received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 853988. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and JDRF INTERNATIONAL.

References I

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011).

Distributed optimization and statistical learning via the alternating direction method of multipliers.

Foundations and Trends® in Machine learning, 3(1):1–122.

Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., Greninger, P., Thompson, I. R., Luo, X., Soares, J., et al. (2012).

Systematic identification of genomic markers of drug sensitivity in cancer cells.

Nature, 483(7391):570–575.

Kim, S. and Xing, E. P. (2012).

Tree-guided group lasso for multi-response regression with structured sparsity, with an application to eqtl mapping.

The Annals of Applied Statistics, 6(3):1095–1117.

References II

Li, Y., Nan, B., and Zhu, J. (2015).

Multivariate sparse group lasso for the multivariate multiple linear regression with an arbitrary group structure.

Biometrics, 71(2):354–363.

Schultheis, B., Carapeti-Marootian, M., Hochhaus, A., Weisser, A., Goldman, J. M., and Melo, J. V. (2002).

Overexpression of SOCS-2 in advanced stages of chronic myeloid leukemia: possible inadequacy of a negative feedback mechanism.

Blood, 99(5):1766–1775.

Tibshirani, R. and Friedman, J. (2020).

A pliable lasso.

Journal of Computational and Graphical Statistics, 29(1):215–225.

THANK YOU