

Shrinkage estimation in regression with categories

Statistical and computational aspects of a new non-convex M-estimator

Benjamin Stokell

joint work with Rajen Shah (University of Cambridge) and Ryan Tibshirani (Carnegie Mellon University)

Statistical Laboratory
Cantab Capital Institute for the Mathematics of Information
Department of Pure Mathematics and Mathematical Statistics
University of Cambridge

b.stokell@statslab.cam.ac.uk

May 24, 2018

- Response $y_1, \dots, y_n \in \mathbb{R}$
- Categorical (factor) variables $j = 1, \dots, p$, each with categories $1, \dots, c_j$
- High-dimensional if $\sum_j (c_j - 1) \gg n$
- Sparsity assumption

- Response $y_1, \dots, y_n \in \mathbb{R}$
- Categorical (factor) variables $j = 1, \dots, p$, each with categories $1, \dots, c_j$
- High-dimensional if $\sum_j (c_j - 1) \gg n$
- Sparsity assumption

Sainsbury's	Waitrose	Lidl	Co-op	M&S	Tesco	Aldi
-------------	----------	------	-------	-----	-------	------

- Response $y_1, \dots, y_n \in \mathbb{R}$
- Categorical (factor) variables $j = 1, \dots, p$, each with categories $1, \dots, c_j$
- High-dimensional if $\sum_j (c_j - 1) \gg n$
- Sparsity assumption

Sainsbury's	Waitrose	Lidl	Co-op	M&S	Tesco	Aldi
-------------	----------	------	-------	-----	-------	------

Range of approaches taking form

$$\hat{\theta} \in \arg \min_{\theta} \sum_{i=1}^n \left(y_i - \hat{\mu} - \sum_{j=1}^p \sum_{k=1}^{c_j} \mathbb{1}_{\{X_{ij}=k\}} \theta_{jk} \right)^2 + \text{penalty}(\theta)$$

- If order/adjacencies known a priori, (generalized) Fused Lasso (Tibshirani et al., 2005)

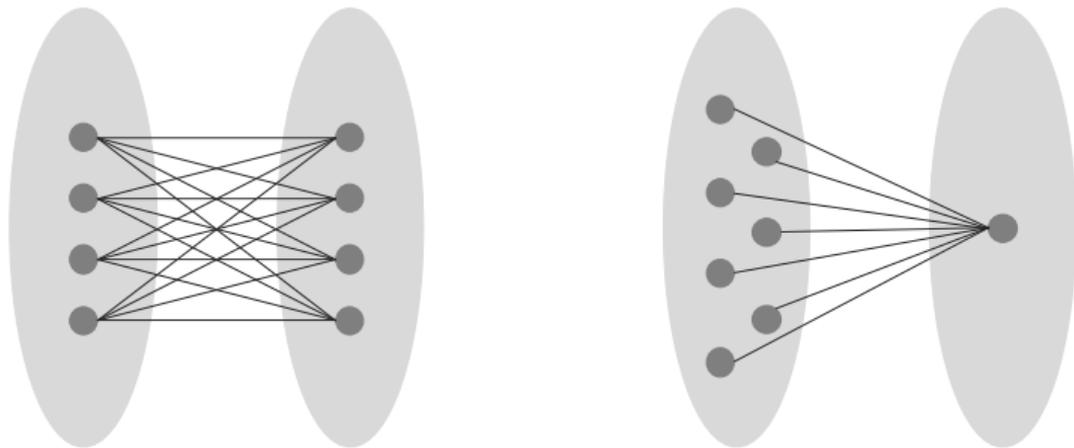
$$\text{penalty}(\theta_j) = \lambda_j \sum_{(k_1 k_2) \in E} |\theta_{jk_1} - \theta_{jk_2}|$$

- If order/adjacencies known a priori, (generalized) Fused Lasso (Tibshirani et al., 2005)

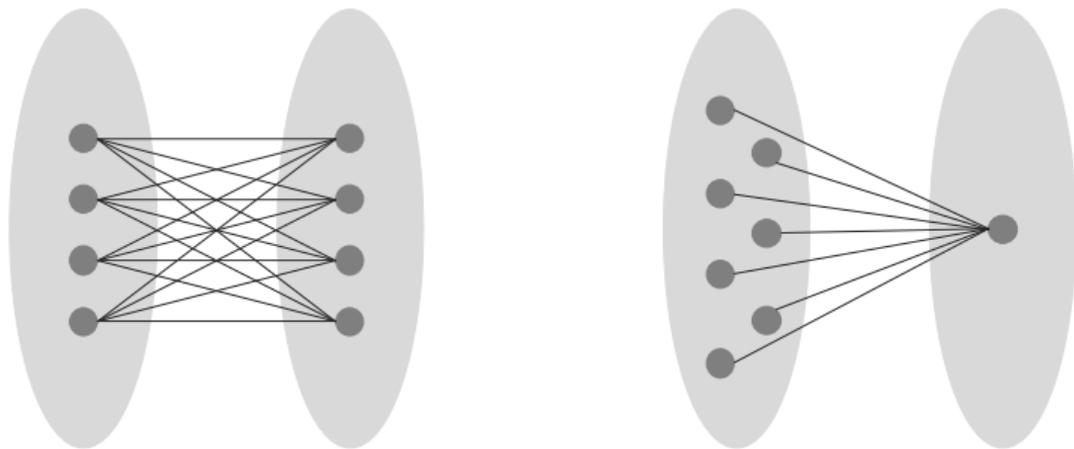
$$\text{penalty}(\theta_j) = \lambda_j \sum_{(k_1 k_2) \in E} |\theta_{jk_1} - \theta_{jk_2}|$$

- Otherwise, have to treat categories symmetrically
 - OSCAR (Bondell & Reich, 2008)
 - Concave pairwise fusion (Ma & Huang, 2017)
 - Many more

Two true levels



Two true levels



Sparse structure of Fused Lasso but without knowledge of ordering?

$$\text{penalty}(\theta_j) = \sum_{k=1}^{c_j-1} \rho(\theta_{j\pi(k+1)} - \theta_{j\pi(k)}) \quad \text{minimizing over } \pi \in \mathcal{S}_{c_j}$$

- Minimizing a nonconvex objective *and* minimizing over permutations

$$\text{penalty}(\theta_j) = \sum_{k=1}^{c_j-1} \rho(\theta_{j\pi(k+1)} - \theta_{j\pi(k)}) \quad \text{minimizing over } \pi \in \mathcal{S}_{c_j}$$

- Minimizing a nonconvex objective *and* minimizing over permutations

Provided ρ is concave, increasing on $[0, \infty)$, symmetric,

- Optimal ordering $\hat{\pi}$ is the ordering of the observations
- Coefficient estimate $\hat{\theta}_j$ also respects this ordering $\hat{\pi}$

- Fast and exact optimization across each variable block through dynamic programming
- Estimation error bounds with minimal assumptions
- Lower bound on probability of recovery of oracle least-squares estimate

Thank you for listening