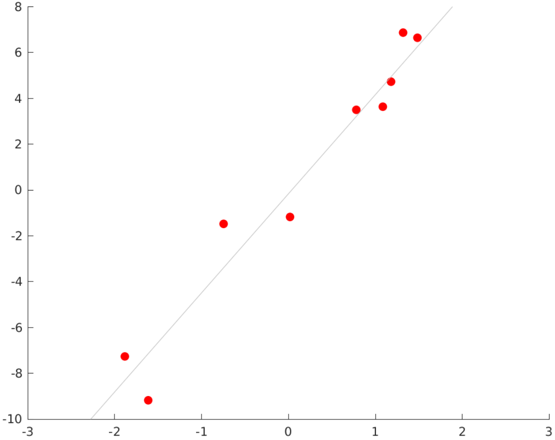
Sparse Mixed Linear Regression with Guarantees: Taming an Intractable Problem with Invex Relaxation

ICML 2022

Adarsh Barik, Dr. Jean Honorio

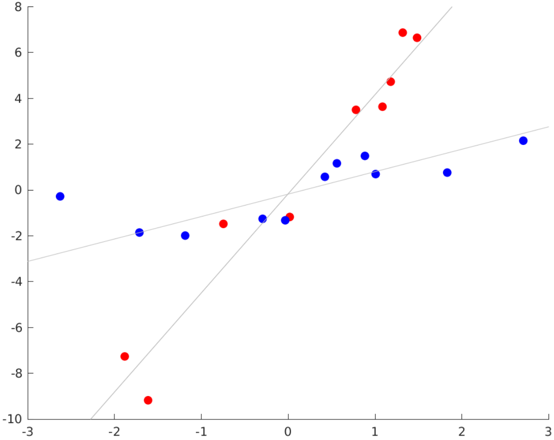
Purdue University

Mixed Linear Regression (MLR)



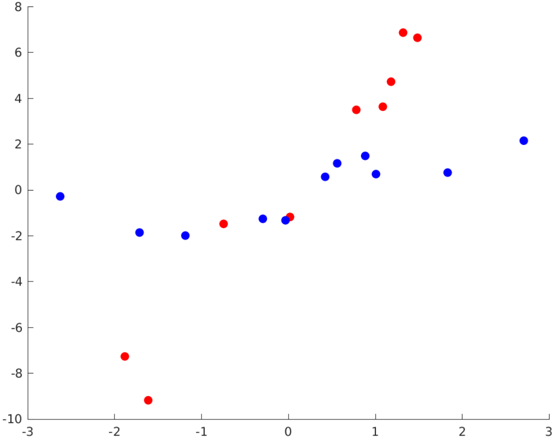
- $y_i = \langle X_i, \beta_1^* \rangle + e_i$

Mixed Linear Regression (MLR)



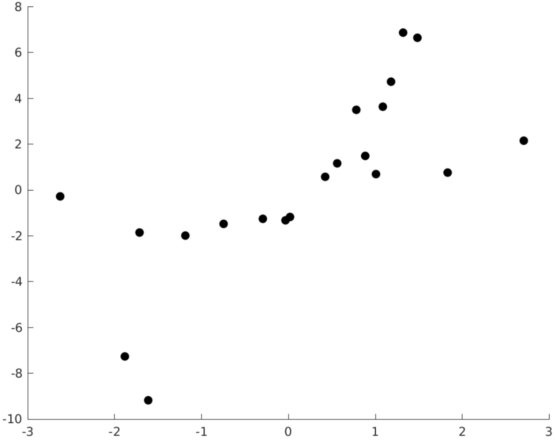
- $y_i = \langle X_i, \beta_1^* \rangle + e_i$
- and $y_i = \langle X_i, \beta_2^* \rangle + e_i$

Mixed Linear Regression (MLR)



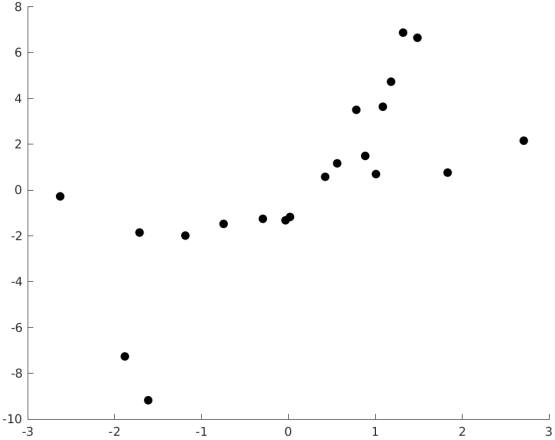
- $y_i = \langle X_i, \beta_1^* \rangle + e_i$
- and $y_i = \langle X_i, \beta_2^* \rangle + e_i$

Mixed Linear Regression (MLR)



- $y_i = z_i^* \langle X_i, \beta_1^* \rangle + (1 - z_i^*) \langle X_i, \beta_2^* \rangle + e_i$
- Recover β_1^*, β_2^* and z_i^* from data $(X_i, y_i), \forall i \in \{1, \cdots, n\}$

Mixed Linear Regression (MLR)



- $y_i = z_i^* \langle X_i, \beta_1^* \rangle + (1 - z_i^*) \langle X_i, \beta_2^* \rangle + e_i$
- Recover β_1^*, β_2^* and z_i^* from data $(X_i, y_i), \forall i \in \{1, \cdots, n\}$
- Applications: Music perception studies [Viele & Tong, 2002], Vehicle merging [Li et al., 2019]

Combinatorial Formulation

- Given attributes $X \in \mathbb{R}^{n \times d}$ and response $y \in \mathbb{R}^n$
- Recover sparse parameters $\beta_1 \in \mathbb{R}^d$ and $\beta_2 \in \mathbb{R}^d$
- Recover hidden labels $z \in \{0, 1\}^n$
- Combinatorial MLR

$$
\begin{array}{ll}
\min_{\beta_1, \beta_2, z} & \frac{1}{n}\left(\sum_{i=1}^n z_i (y_i - \langle X_i, \beta_1 \rangle)^2 + (1 - z_i)(y_i - \langle X_i, \beta_2 \rangle)^2\right) \\
\text{subject to} & z_i \in \{0, 1\}, \quad \forall i \in \{1, \cdots, n\} \\
& \|\beta_1\|_0 = s_1, \quad \|\beta_2\|_0 = s_2
\end{array}
$$

Combinatorial Formulation

- Given attributes $X \in \mathbb{R}^{n \times d}$ and response $y \in \mathbb{R}^n$
- Recover sparse parameters $\beta_1 \in \mathbb{R}^d$ and $\beta_2 \in \mathbb{R}^d$
- Recover hidden labels $z \in \{0, 1\}^n$
- Combinatorial MLR

$$\begin{aligned} \min_{\beta_1, \beta_2, z} \quad & \frac{1}{n}\left(\sum_{i=1}^{n} z_i (y_i - \langle X_i, \beta_1 \rangle)^2 + (1 - z_i)(y_i - \langle X_i, \beta_2 \rangle)^2\right) \\ \text{subject to} \quad & z_i \in \{0, 1\}, \quad \forall i \in \{1, \cdots, n\} \\ & \|\beta_1\|_0 = s_1, \quad \|\beta_2\|_0 = s_2 \end{aligned}$$

Non-convex even after continuous relaxation and regularization!

Invex Relaxation

- Given attributes $X \in \mathbb{R}^{n \times d}$ and response $y \in \mathbb{R}^n$ and $S_i = \begin{bmatrix} X_i \\ -y_i \end{bmatrix} \begin{bmatrix} X_i^T & -y_i \end{bmatrix}$
- Recover sparse matrices $W \in \mathbb{R}^{d+1 \times d+1}$, $U \in \mathbb{R}^{d+1 \times d+1}$ and hidden labels $t \in \{-1, +1\}^n$
- Invex MLR

$$
\begin{array}{ll}
\min_{t, W, U} & \frac{1}{2}\left(\sum_{i=1}^n \langle S_i, W + U \rangle + t_i \langle S_i, W - U \rangle\right) + \lambda_1 \|W(:)\|_1 + \lambda_2 \|U(:)\|_1 \\
\text{subject to} & W \succeq \mathbf{0}, \quad U \succeq \mathbf{0} \\
& W_{d+1,d+1} = 1, \quad U_{d+1,d+1} = 1 \\
& \|t\|_\infty \leqslant 1
\end{array}
$$

Intuition: $\begin{bmatrix} \beta_1 \\ 1 \end{bmatrix} \begin{bmatrix} \beta_1^T & 1 \end{bmatrix} \succeq \mathbf{0}$, $\begin{bmatrix} \beta_2 \\ 1 \end{bmatrix} \begin{bmatrix} \beta_2^T & 1 \end{bmatrix} \succeq \mathbf{0}$, $z_i = \frac{t_i + 1}{2}$

Invexity

- A function ϕ is η-invex iff

$$\phi(u) - \phi(v) \geqslant \nabla\phi(v)^\intercal \eta(u, v) \quad \forall u, v$$

- A convex function ϕ is η-invex for $\eta(u, v) = u - v$.
- KKT conditions
  - necessary even for non-convex problems
  - sufficient for invex problems [Hanson, 1981]
  - Come up with unique primal and dual variables which satisfy KKT conditions

Invex Formulation

- Invex MLR

$$
\begin{aligned}
\min_{t,W,U} \quad & \tfrac{1}{2}\left(\textstyle\sum_{i=1}^{n}\langle S_i, W+U\rangle + t_i\langle S_i, W-U\rangle\right) + \lambda_1\|W(:)\|_1 + \lambda_2\|U(:)\|_1 \\
\text{subject to} \quad & W \succeq \mathbf{0}, \quad U \succeq \mathbf{0} \\
& W_{d+1,d+1} = 1, \quad U_{d+1,d+1} = 1 \\
& \|t\|_\infty \leqslant 1
\end{aligned}
$$

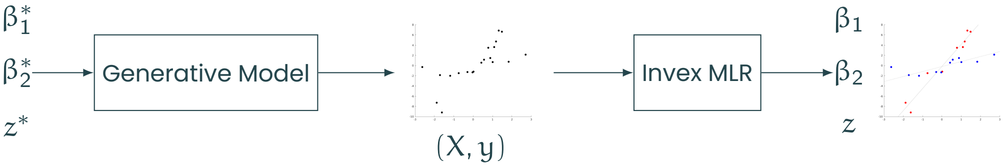
- Functions are η-invex for

$$
\eta((t, W, U), (\widetilde{t}, \widetilde{W}, \widetilde{U})) = \begin{bmatrix} \mathbf{0} \\ -\widetilde{W} \\ -\widetilde{U} \end{bmatrix}
$$

Generative Model

- True sparse parameters and hidden labels: $\beta_1^* \in \mathbb{R}^d, \quad \beta_2^* \in \mathbb{R}^d, \quad z^* \in \{0,1\}^n$
- Random attributes and independent noise: $X \in \mathbb{R}^{n \times d}, \quad e \in \mathbb{R}^n$
- The response $y_i \in \mathbb{R}$ is generated by:

$$y_i = z_i^* \langle X_i, \beta_1^* \rangle + (1 - z_i^*) \langle X_i, \beta_2^* \rangle + e_i \quad \forall i \in \{1, \cdots, n\}$$



Exact Recovery Guarantee

- Let $S(\beta) = \{i | \beta_i \neq 0\}$ be the support of $\beta \in \mathbb{R}^d$
- $\sum_{i=1}^{n} z_i^* = n_1, \quad n_2 = n - n_1$
- Invex MLR outputs the true solution

$$S(\beta_1) = S(\beta_1^*), \quad S(\beta_2) = S(\beta_2^*), \quad z = z^*$$

with high probability provided that:

$$n_1 \geqslant \Omega(|S(\beta_1^*)|^3 \log d), \quad n_2 \geqslant \Omega(|S(\beta_2^*)|^3 \log d)$$

- Furthermore, $\|\beta_1 - \beta_1^*\|_2 \leqslant \delta, \quad \|\beta_2 - \beta_2^*\|_2 \leqslant \delta.$

Thank You!