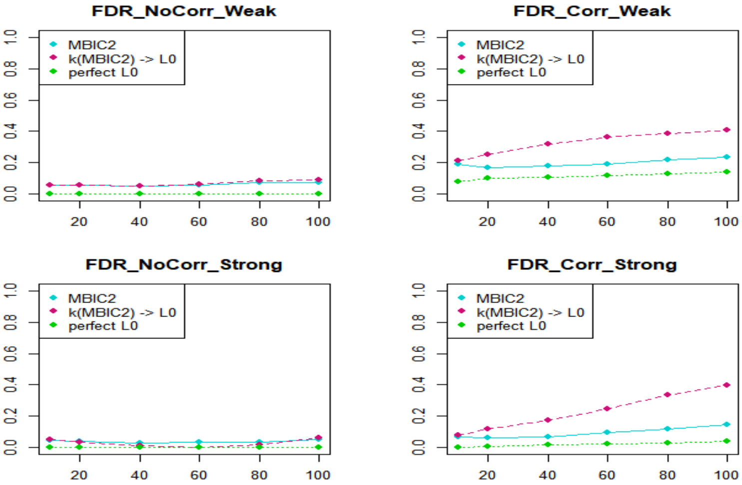
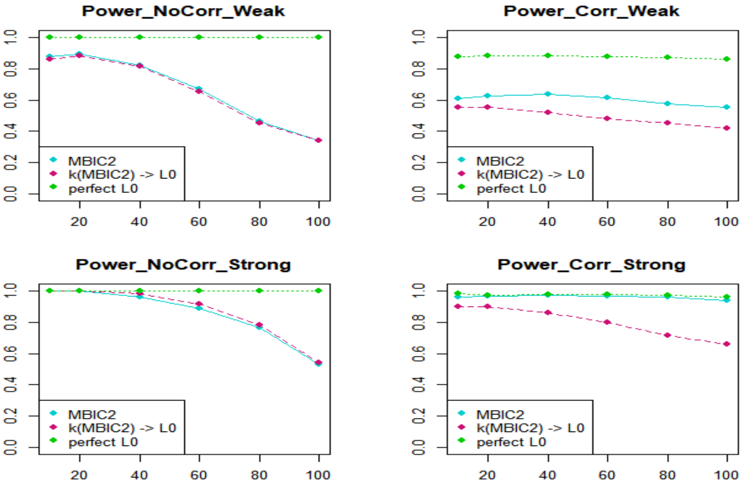
Inne metody


FDR_NoCorr_Weak
FDR_Corr_Weak
FDR_NoCorr_Strong
FDR_Corr_Strong
MBIC2
k(MBIC2) -> L0
perfect L0
0.0
0.2
0.4
0.6
0.8
1.0
20
40
60
80
100

Inne metody c.d.



Problem wyboru optymalnego podzbioru (BSS)

Dane: $\mathrm{y}_{n\times 1}, \mathrm{X}_{n\times p}, k$

$$\min_{\boldsymbol{\beta}} \frac{1}{2}||\mathrm{y} - \mathrm{X}\boldsymbol{\beta}||_2^2 \text{ takie, że } ||\boldsymbol{\beta}||_0 \leq k$$

Mieszana całkowitoliczbowa optymalizacja kwadratowa

Mieszana całkowitoliczbowa optymalizacja kwadratowa (ang. mixed integer quadratic optimization - MIO) jest zbiorem problemów opisanych następującym wzorem:

$$\min \ \alpha^T Q\alpha + \alpha^T a$$

$$\text{takie, że} \quad A\alpha \leq b$$

$$\alpha_i \in \{0, 1\}, \quad i \in \mathcal{I},$$

$$\alpha_j \geq 0, \quad j \notin \mathcal{I},$$

gdzie $a \in \mathbb{R}^m$, $A \in \mathbb{R}^{k\times m}$, $b \in \mathbb{R}^k$ oraz $Q \in \mathbb{R}^{m\times m}$ nieujemnie określona są danymi parametrami. Wektor $\alpha \in \mathbb{R}^m$ zawiera wartości dyskretne dla α_i, $i \in \mathcal{I}$ i ciągłe dla $\alpha \notin \mathcal{I}$, przy $\mathcal{I} \subset \{1, ..., m\}$.

Mieszana całkowitoliczbowa optymalizacja kwadratowa c.d.

Niektórymi podproblamami MIO są problemy kwadratowej optymalizacji wypukłej ($\mathcal{I} = \emptyset$), mieszane całkowitoliczbowe i liniowe problemy optymalizacyjne ($Q = 0_{m \times m}$). Przykładowe zaawansowane programy analizujące te problemy to CPLEX, Gurobi i Xpress.

Sposoby wyrażenia problemu optymalnego podzbioru przez MIO

Proste sformułowanie problemu BSS może być następujące:

$$Z_1 = \min_{\boldsymbol{\beta},z} \frac{1}{2}||\mathrm{y} - \mathrm{X}\boldsymbol{\beta}||_2^2$$

$$\text{takie, że:} \quad -\mathcal{M}_U z_i \leq \beta_i \leq \mathcal{M}_U z_i, \quad i = 1, ..., p,$$

$$z_i \in \{0, 1\}$$

$$\sum_{i=1}^{p} z_i \leq k,$$

gdzie $\mathrm{z} \in \{0,1\}^p$ oraz $\mathcal{M}_U$ jest stałą taką, że rozwiązanie $\hat{\boldsymbol{\beta}}$ spełnia $||\hat{\boldsymbol{\beta}}||_\infty \leq \mathcal{M}_U$. Jeśli $z_i = 1$, to $|\beta_i| \leq \mathcal{M}_U$, jeśli $z_i = 0$, to $\beta_i = 0$, skąd $\sum_{i=1}^{p} z_i$ jest ilością niezerowych współrzędnych w wektorze $\hat{\boldsymbol{\beta}}$.

Sposoby wyrażenia problemu optymalnego podzbioru przez MIO c.d.

Powyższe sformułowanie jest powiązane z problemem LASSO. Rozważmy uwypuklenie zbioru rozwiązań z ich ograniczeniami

$$\text{Conv}\left(\left\{\boldsymbol{\beta} : |\beta_i| \leq \mathcal{M}_U z_i, z_i \in \{0,1\},\ i = 1, ..., p,\ \sum_{i=1}^{p} z_i \leq k\right\}\right)$$

$$= \{\boldsymbol{\beta} : ||\boldsymbol{\beta}||_\infty \leq \mathcal{M}_U,\ ||\boldsymbol{\beta}||_1 \leq k\mathcal{M}_U\} \subseteq \{\boldsymbol{\beta} : ||\boldsymbol{\beta}||_1 \leq k\mathcal{M}_U\}.$$

Sposoby wyrażenia problemu optymalnego podzbioru przez MIO c.d.

$$Z_2 = \min_{\boldsymbol{\beta}} \frac{1}{2}||\mathsf{y} - \mathsf{X}\boldsymbol{\beta}||_2^2 \text{ takie, że } ||\boldsymbol{\beta}||_\infty \leq \mathcal{M}_U, \ ||\boldsymbol{\beta}||_1 \leq k\mathcal{M}_U,$$

$$Z_3 = \min_{\boldsymbol{\beta}} \frac{1}{2}||\mathsf{y} - \mathsf{X}\boldsymbol{\beta}||_2^2 \text{ takie, że } ||\boldsymbol{\beta}||_1 \leq k\mathcal{M}_U,$$

gdzie Z_3 jest dokładnie problemem LASSO. Jest to uogólnienie problemu Z_2, w którym dodatkowo kontrolujemy poszczególne wartości β_i. W związku z tym zachodzą następujące nierówności $Z_3 \leq Z_2 \leq Z_1$, gdzie w praktyce są to nierówności ostre. Okazuje się to istotne w kontekście szacowania rozwiązania przez MIO, ponieważ programy najpierw wykorzystują ciągłe uogólnienie Z_1.

Sposoby wyrażenia problemu optymalnego podzbioru przez MIO c.d.

Będziemy rozważali również następujące sformułowanie:

$$Z_4 = \min_{\boldsymbol{\beta}} \frac{1}{2}||\mathsf{y} - \mathsf{X}\boldsymbol{\beta}||_2^2$$

$$\text{takie, że:} \quad ||\boldsymbol{\beta}||_0 \leq k,$$

$$||\boldsymbol{\beta}||_\infty \leq \mathcal{M}_U, \quad ||\boldsymbol{\beta}||_1 \leq \mathcal{M}_\ell,$$

$$||\mathsf{X}\boldsymbol{\beta}||_\infty \leq \mathcal{M}_U^\xi, \quad ||\mathsf{X}\boldsymbol{\beta}||_1 \leq \mathcal{M}_\ell^\xi.$$

Sposoby wyrażenia problemu optymalnego podzbioru przez MIO c.d.

Dla dostatecznie dużej wartości stałej $\mathcal{M}_U$ dla sformułowania Z_1 lub stałych $\mathcal{M}_U, \mathcal{M}_\ell, \mathcal{M}_U^\xi, \mathcal{M}_\ell^\xi$, dla sformułowania Z_4 rozwiązanie problemu będzie rozwiązaniem wyjściowego problemu. Kiedy ich wartość będzie zbyt mała, możemy otrzymać rozwiązanie inne niż dla BSS. Pokażemy teraz jak wyliczyć te ograniczenia z analizowanych danych.

Specyfikacja parametrów przez spójność i ograniczone wartości własne

Definiujemy funkcję kumulowanej spójności

$$\mu[k] = \max_{|I|=k} \max_{j \notin I} \sum_{i \in I} |\langle X_j, X_i \rangle|,$$

gdzie $X_j, j = 1, ..., p$ są kolumnami macierzy X. W szczególności dla $k = 1$ otrzymujemy spójność jako miarę maksymalnej korelacji między kolumnami macierzy X: $\mu = \mu[1] = \max_{i \neq j} |\langle X_j, X_i \rangle|$.

Specyfikacja parametrów przez spójność i ograniczone wartości własne c.d.

Skorzystamy również z następującego wyrażenia warunku ograniczonych wartości własnych:

$$\lambda_{min}(X_I^T X_I) \geq \gamma_k \text{ dla dowolnego } I \subset \{1, ..., p\} \text{ takiego, że } |I| \leq k,$$

gdzie $\lambda_{min}(X_I^T X_I)$ jest najmniejszą wartością własną macierzy $X_I^T X_I$.

Specyfikacja parametrów przez spójność i ograniczone wartości własne c.d.

Twierdzenie

Mamy następujące ograniczenia:
(a) $\mu[k] \leq \mu k$,
(b) $\gamma_k \geq 1 - \mu[k-1] \geq 1 - \mu(k-1)$.

Specyfikacja parametrów przez spójność i ograniczone wartości własne c.d.

Użyteczne będzie również pojęcie normy operatorowej macierzy. Dla ustalonych (p, q) normę operatorową macierzy A definiujemy jako $||\mathrm{A}|| := \max_{||\mathrm{u}||_q=1} ||\mathrm{Au}||_p$. Będziemy używali jedynie $(1, 1)$ normy operatorowej.

Specyfikacja parametrów przez spójność i ograniczone wartości własne c.d.

Twierdzenie

Niech kolumny macierzy X *mają normę* ℓ_2 *równą 1. Dla dowolnego* $I \subset \{1, ..., p\}$, *gdzie* $|I| = k$, *mamy:*
(a) $||\mathrm{X}_I^T \mathrm{X}_I - \mathrm{I}||_{1,1} \leq \mu[k-1]$,
(b) Jeśli $\mathrm{X}_I^T \mathrm{X}_I$ *jest odwracalna oraz* $||\mathrm{X}_I^T \mathrm{X}_I - \mathrm{I}||_{1,1} \leq 1$, *to* $||(\mathrm{X}_I^T \mathrm{X}_I)^{-1}||_{1,1} \leq \frac{1}{1-\mu[k-1]}$.

Specyfikacja parametrów przez spójność i ograniczone wartości własne c.d.

Pamiętając, że $X_j, j = 1, ..., p$ oznaczają kolumny X oraz że $||X_j||_2 = 1$, wprowadźmy sortowanie korelacji:

$$|\langle X_{(1)}, y\rangle| \geq |\langle X_{(2)}, y\rangle| \geq ... \geq |\langle X_{(p)}, y\rangle|$$

Przez $x_i, i = 1, ..., n$ będziemy oznaczali wiersze macierzy X. Dodatkowo, niech $||x_i||_{1:k}$ oznacza sumę k największych co do wartości bezwzględnej elementów $x_{ij}, j = 1, ..., p$.
Wówczas dla dowolnego $k \geq 1$ takiego, że $\mu[k-1] < 1$ rozwiązanie $\hat{\beta}$ problemu BSS spełnia następujące ograniczenia:

Twierdzenie

- $$||\hat{\beta}||_1 \leq \frac{1}{1-\mu[k-1]} \sum_{j=1}^{k} |\langle X_{(j)}, y\rangle|,$$
- $$||\hat{\beta}||_\infty \leq \min\left\{ \frac{1}{\gamma_k}\sqrt{\sum_{j=1}^{k} |\langle X_{(j)}, y\rangle|^2}, \frac{1}{\sqrt{\gamma_k}}||y||_2 \right\},$$
- $$||X\hat{\beta}||_1 \leq \min\left\{ \sum_{i=1}^{n} ||x_i||_\infty ||\hat{\beta}||_1, \sqrt{k}||y||_2 \right\},$$
- $$||X\hat{\beta}||_\infty \leq \left(\max_{i=1,\dots,n} ||x_i||_{1:k} \right) ||\hat{\beta}||_\infty.$$

L0Learn

`L0Learn` is a fast toolkit for L0-regularized learning. L0 regularization selects the best subset of features and can outperform commonly used feature selection methods (e.g., L1 and MCP) under many sparse learning regimes. The toolkit can (approximately) solve the following three problems

$$\min_{\beta_0,\beta} \sum_{i=1}^{n} \ell(y_i, \beta_0 + \langle x_i, \beta \rangle) + \lambda ||\beta||_0 \qquad (L0)$$

$$\min_{\beta_0,\beta} \sum_{i=1}^{n} \ell(y_i, \beta_0 + \langle x_i, \beta \rangle) + \lambda ||\beta||_0 + \gamma ||\beta||_1 \quad (L0L1)$$

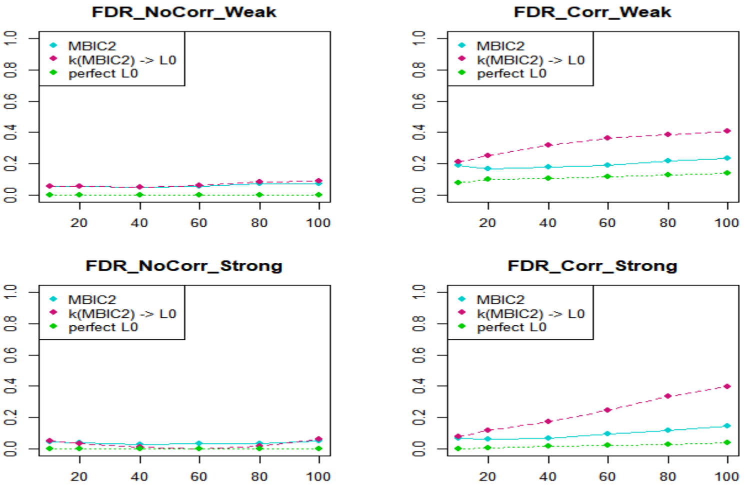
$$\min_{\beta_0,\beta} \sum_{i=1}^{n} \ell(y_i, \beta_0 + \langle x_i, \beta \rangle) + \lambda ||\beta||_0 + \gamma ||\beta||_2^2 \quad (L0L2)$$

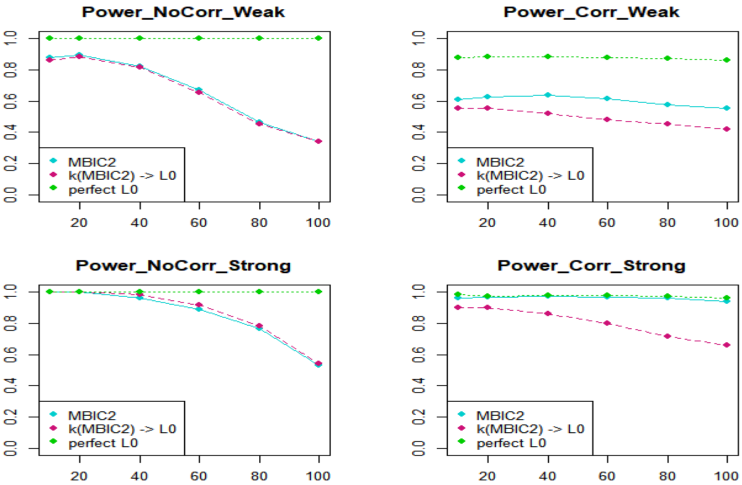
Przykłady

Porównanie z kryteriami informacyjnymi

W symulacji porównujemy następujące metody:

- mBIC2 z procedurą stepwise z biblioteki *bigstep*
- *L0Learn.fit* z karą L0 i optymalną rzadkością wektora współczynników
- *L0Learn.fit* z karą L0 i rzadkością wektora współczynników najbliższą tej, wybranej przez mBIC2.


FDR_NoCorr_Weak
MBIC2
k(MBIC2) -> L0
perfect L0
FDR_Corr_Weak
MBIC2
k(MBIC2) -> L0
perfect L0
FDR_NoCorr_Strong
MBIC2
k(MBIC2) -> L0
perfect L0
FDR_Corr_Strong
MBIC2
k(MBIC2) -> L0
perfect L0



Power_NoCorr_Weak
Power_Corr_Weak
Power_NoCorr_Strong
Power_Corr_Strong
0.0
0.2
0.4
0.6
0.8
1.0
20
40
60
80
100
MBIC2
k(MBIC2) -> L0
perfect L0