Model selection with stepdown procedures and thresholded Lasso for high dimensional linear regression

Konrad Furmańczyk

Faculty of Applied Infomatics and Mathematics, Warsaw University of Life Sciences (SGGW)

Będlewo 1.12.2016

Outline

- Linear model with Subgaussian errors;
- Two steps selection procedure TLSD;
- Stepdown procedure;
- The consistency of the selection procedure;
- Simulation study.

$$\mathbb{Y} = \mathbb{X}\beta + \varepsilon,$$

(ϵ_i) i.i.d., $E(\epsilon_1) = 0$, $Var(\epsilon_1) = \sigma^2$,
ϵ_1 is Subgaussian with a constant $\sigma > 0$ i.e.
$E(exp(u\epsilon_1)) \leq exp(\sigma^2 u^2/2)$ for all $u \in R$
$\mathbb{X}$ -deterministic matrix of dimension n x p; $\beta \in R^p$;
$p = p(n) > > n$;

We will consider only sparse models (finite number of β_i is different from 0) and matrix $\mathbb{X}$ satisfies some regular conditions.

LASSO
$\widehat{\beta_L} = argmin_{\beta \in R^p} \left(\frac{1}{2n} \parallel \mathbb{Y} - \mathbb{X}\beta \parallel_2^2 + \lambda_n \parallel \beta \parallel_1 \right)$
for some $\lambda_n > 0$

Thresholded LASSO (TL)
$\widehat{\beta_{TL,j}} = \widehat{\beta_{L,j}} \mathbf{1} \left\{ \left| \widehat{\beta_{L,j}} \right| \geq \delta_n \right\}$ for $j = 1, ..., p$
for some threshold $\delta_n > 0$;

Sparse Model

Sparse Model

$I_0 = \{j : \beta_j \neq 0\}$,
$I_1 = \{j : \beta_j = 0\}$
$|I_0| = p_0 < n$
p_0 -is constant, does not depend on n

Two steps selection procedure TLSD

At the first step (TL)
we choose a set of variables $S_1 = \left\{1 \leq j \leq p : \left|\widehat{\beta_{L,j}}\right| \geq \delta_n\right\}$

At the second step we use a stepdown procedure (SD)
of multitesting

(h_0) $\quad$ H_i: $\beta_i = 0$ vs. H_i': $\beta_i \neq 0$ dla $i \in S_1$

Stepdown procedure

Define p-value for hypothesis testing H_i versus H_i' as $\pi_i = 2\left(1 - \Phi(|t_i|)\right)$ for $i \in S_1$, where Φ is the distribution function of $N\left(0, 1\right)$.

Test Statistics $t_i = \widehat{\beta_{ols,i}} / se\left(\widehat{\beta_{ols,i}}\right)$, where

$$\widehat{\beta_{ols}} = \left(\mathbb{X}_{S1}'\mathbb{X}_{S1}\right)^{-1}\mathbb{X}_{S1}'\mathbb{Y}$$

$$se\left(\widehat{\beta_{ols,i}}\right) = \begin{cases} \sigma\sqrt{m_{i,i}} & \text{if } \sigma \text{ is known} \\ \mathrm{S}\sqrt{m_{i,i}} & \text{if } \sigma \text{ unknown} \end{cases}$$

S -some consistent estimator of σ and $\left(\mathbb{X}_{S1}'\mathbb{X}_{S1}\right)^{-1} = \left(m_{i,j}\right)_{i,j \in S_1}$

Stepdown procedure

Let $s_1 = |S_1|$
We have p-values ordered $\pi_{(1)} \leq ... \leq \pi_{(s_1)}$ and respectively null hypotheses $H_{(1)} \leq ... \leq H_{(s_1)}$

$\alpha_1 \leq ... \leq \alpha_{s1}$ -some constants (may depend on n)

If $\pi_{(1)} > \alpha_1$, then we do not reject any hypothesis H_i;
(h_1) otherwise if $\pi_{(1)} \leq \alpha_1, ..., \pi_{(r)} \leq \alpha_r$
then, we reject $H_{(1)}, ..., H_{(r)}$, where r is the largest number satisfying (h_1) .

Examples of SD procedures

a) *Holm* $\alpha_j = \frac{q_n}{s_1+1-j}$

b) *UHolm* $\alpha_j = \frac{([\gamma j]+1)q_n}{s_1+[\gamma j]+1-j}$ dla $0 \le \gamma \le 1$

c) *BH* $\alpha_j = \frac{jq_n}{s_1}$

d) *Bonferroni* $\alpha_j = \frac{q_n}{s_1}$ for some $q_n \to 0$

(A_1) $\quad \alpha_{s1} \to 0$ as $n \to \infty$

(A_2) $\quad (1/n)\log(1/\alpha_1) \to 0$ as $n \to \infty$

Remark. For $q_n = 1/(n\log(n))$ the conditions $(A_1) - (A_2)$ are satisfied.

Regular conditions for a matrix $\mathbb{X}$

(B_1) $\| x_j \|_2 / \sqrt{n} = 1$ for $j = 1, ..., p$, where x_j -jth column of matrix $\mathbb{X}$

(B_2) *let* $\mathrm{v}_{I1} = \{v_i : i \in I_1\}$, $v_{I0} = \{v_i : i \in I_0\}$;
$C(I_0; 3) = \{v \in R^p : \| v_{I1} \|_1 \leq 3 \| v_{I0} \|_1\}$;
for some $\gamma > 0$ the below condition is satisfied
$(1/n)\, v' \mathbb{X}' \mathbb{X} v \geq \gamma \| v \|_2^2$ for every $v \in C(I_0; 3)$

Regular conditions

regular conditions for model and estimators

(B_3) $\min_{j \in I_0} |\beta_j| \geq C_1 \lambda_n$ for some constant $C_1 > 0$

(B_4) $C_2 \lambda_n \leq \delta_n \leq \lambda_n (C_1 - 3/\gamma)$ for some constant $C_2 > 0$

(B_5) $\lambda_n = C_\lambda \sqrt{\log(p)/n}$ for $C_\lambda = 2\sigma$.

consistency condition for S

(C) $\mathrm{S} \to^P \sigma$ as $n \to \infty$

The consistency of selection procedure

Consistency Theorem

Any TLSD procedure satysfying: $(A_1) - (A_2), (B_1) - (B_5)$ if σ is known and additionally (C) if σ is unknown, is consistent for selection problem in linear model.

The consistency of selection procedure

A selection procedure is consistent if $P\left(\widehat{I} = I_0\right) \to 1$ as $n \to \infty$, where $\widehat{I}$ is the number of chosen significant variables.

Auxilliary results

Lemma 1

Let (B_2), (B_4), (B_5) be satisfied.
Then for every $r > 0$ at least with probability $1 - 2\exp(-\frac{1}{2}(r-2)\log(p))$, we have $|S_1| \leq p_0\left(1 + C_3/\gamma^2\right)$ for some constant $C_3 > 0$.

Lemma 2

Let (B_2), (B_3), (B_5) be satisfied.
Then for every $r > 0$ at least with probability $1 - 2\exp(-\frac{1}{2}(r-2)\log(p))$, we have $I_0 \subset S_1$.

Sketch of the proof

A selection procedure is consistent if $P(V \geq 1) \to 0$ and $P(R \neq p_0) \to 0$ as $n \to \infty$, where V -the number of false rejected predoctors, R -the number of all rejected predictors using TLSD procedure. From Lemma 2 it is sufficient to show that $P\left(\widetilde{V} \geq 1\right) \to 0$ and $P\left(\widetilde{R} \neq p_0\right) \to 0$, where $\widetilde{V}$ -the number of false rejected predictors, $\widetilde{R}$ -the number of all rejected predictors using the stepdown procedure $(h_0) - (h_1)$.

Sketch of the proof

By Lemma 1 it is sufficient to show (*p. Furmanczyk*, 2016)

(i) $\mathrm{P}(\pi_i \leq \alpha_{s1}) \to 0$ for $i \in I_1$

(ii) $\max_{j \in I_0} (1 - F_j(\alpha_1)) \in 0$ as $n \to \infty$, where F_j -distribution function of p-value π_j for $j \in I_0$.

The conditions $(i) - (ii)$
are implied by the conditions $(A_1), (A_2), (C)$.

Simulations

The matrix $\mathbb{X}$ is fixed for all simulations generated from $N_p(0, Id)$ and normalized

Simulated models

$(M1)$ $\mathrm{Y}=\sum_{j=1}^{p_0} 2X_j + \epsilon$

$(M2)$ $\mathrm{Y}=1.3\mathrm{X}_1 + X_2 + 1.4X_3 + 1.5X_4 + \epsilon$

$(M3)$ $\mathrm{Y}=1.3\mathrm{X}_1 + 1.3X_2 + X_3 + X_4 + 1.4X_5 + 1.4X_6 + 1.5X_7 + 1.5X_8 + \epsilon,$

where $\epsilon \sim N(0,1)$

In simulation we set $\delta_n = \sqrt{\log(p)/n}$ and
$\lambda_n = 2\sqrt{\log(p)/n} * seq(10, 0.05, by = -0.05)$ for cross-validation for Lasso.
1000 MC replications of model selection
$(M1) - (M3)$ were carried out.
We used the procedur TLSD(Holm, UHolm, Bonf., BH) and for comaparision TL, SCAD.

SCAD (Smoothly Clipped Absolute Deviation - Fan and Li (2001))

$$\widehat{\beta} = argmin_{\beta \in R^p} \parallel \mathbb{Y} - \mathbb{X}\beta \parallel_2^2 /n + \sum_{i=1}^{p} J_\lambda\left(|\beta_i|\right)$$

The penalty for $a = 3.7$

$$J_{\lambda}(\theta) = \begin{cases} \lambda|\theta| & \text{for } |\theta| \leq \lambda \\ -\left(\theta^2 - 2a\lambda|\theta| + \lambda^2\right)/\left(2(a-1)\right) & \text{for } \lambda < |\theta| \leq a\lambda \\ (a+1)\lambda^2/2 & \text{for } |\theta| > a\lambda \end{cases}$$

for LASSO we have the penalty

$$J_{\lambda}(\theta) = \lambda|\theta|$$

Results for M1 simulations

	p=500 p_0=5	p=500 p_0=10	p=500 p_0=20	p=1000 p_0=20	p=2000 p_0=5	p=2000 p_0=10
Bonf	987	986	983	980	995	990
Holm	980	965	904	904	994	979
UHolm_0.01	980	965	904	904	994	979
UHolm_0.1	980	964	895	899	994	979
UHolm_0.5	980	964	890	895	994	979
UHolm_0.9	980	965	890	895	994	979
BH	980	963	880	889	994	978
SCAD	13	26	43	33	3	6
TL	974	930	589	573	994	960

Table 1. Frequencies of the true model that are selected in 1000 simulations for $n = 200$.

Results for M2 simulations

	p=500 n=100	p=500 n=200	p=1000 n=100	p=1000 n=200	p=2000 n=100	p=2000 n=200
Bonf	991	993	995	992	997	997
Holm	989	990	994	991	997	996
UHolm_0.01	989	990	994	991	997	996
UHolm_0.1	989	990	994	991	997	996
UHolm_0.5	989	990	994	991	997	996
UHolm_0.9	989	990	994	991	997	996
BH	988	988	994	991	997	996
SCAD	38	14	33	7	17	7
TL	986	985	993	991	996	995

Table 2. Frequencies of the true model that are selected in 1000 simulations

Results for M3 simulations

	p=500 n=100	p=500 n=200	p=1000 n=100	p=1000 n=200	p=2000 n=100	p=2000 n=200
Bonf	977	988	974	992	984	990
Holm	929	972	951	989	964	987
UHolm_0.01	929	972	951	989	964	987
UHolm_0.1	929	972	951	989	964	987
UHolm_0.5	929	972	951	989	963	987
UHolm_0.9	929	972	950	989	963	987
BH	924	970	945	989	960	987
SCAD	68	15	43	10	30	4
TL	833	957	889	987	897	985

Table 3. Frequencies of the true model that are selected in 1000 simulations for

References

Benjamini, Y., Liu, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *J. Statist. Plann. Infer.* **82**, 163-170.

Bunea, F., Wegkamp, M.H., Auguste, A. (2006). Consistent variable selection in high dimensional regression via multiple testing. *J. Statist. Plann. Infer.* **136**, 12, 4349-4364.

Fan, J.and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96**, 1348–1360.

Friedman, J., Hastie, T., Simon, N. and Tibshirani, R. (2015), glmnet: Lasso and elastic-net regularized generalized linear models. R package version 2.0.

Furmańczyk, K. (2014). Selection in parametric models via some stepdown procedures. *Appl. Math. (Warsaw)* **41**, 81-92.

- Furmańczyk, K. (2015). On some stepdown procedures with application to consistent variable selection in linear regression. *Statistics* **49**, 614-628.
- Furmańczyk, K. (2016). Variable selection using stepdown procedures in high-dimensional linear models. *Appl. Math. (Warsaw)* DOI: 10.4064/am2286-6-2016
- Furmańczyk, K. (2016). Stepdown procedures with thresholded Lasso for high dimensional parametric model selection. preprint
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* **6**, 65-70.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267-288.