

Raport 3 cz.2

Aleksander Milach

5 June 2019

Raport ma dwie części i jest rozdzielony w zadaniu 3. Na końcu pierwszej części jest wynik pojedynczego eksperymentu z zadania 3. Poniżej są uśrednione wyniki 20 eksperymentów z zadania 3. Z niewiadomych mi powodów po przekroczeniu pewnej ilości linijek Markdown uniemożliwił efektywne obliczanie kolejnych (prawidłowo liczących w R) chunków i wyrzucał pewien niezwiązany z kodem błąd; przerzucenie tego samego kodu do nowego pliku rozwiązywało problem (przyp. aut.).

Zadanie 3 cd.

	Bbetahat	SEbetahat	SEphat	T	F	FDP
GLM	0.0026982	5.2040904	58.350504	4.75	24.60	0.8293337
BIC	-0.0005781	0.2019573	2.309311	4.90	2.00	0.2576299
mBIC	-0.0013258	0.1499382	5.127683	4.10	0.05	0.0083333
mBIC2	-0.0012079	0.1741653	3.753719	4.45	0.30	0.0575000
LASSO	-0.0006317	0.2211153	3.461604	4.85	7.10	0.5533186
SLOPE	-0.0005672	0.3364746	7.681158	5.00	29.95	0.8383848

Ze względu na dodanie korelacji otrzymujemy gorsze wyniki. Funkcja glm wciąż ma największe błędy estymacji. glm i SLOPE dodają bardzo dużo fałszywych odkryć, LASSO trochę mniej. Najlepiej poradziły sobie kryteria informacyjne: mBIC miał zdecydowanie najmniejsze FDP kosztem zmniejszonej wykrywalności sygnału, a mBIC2 najczęściej znajdował sygnał przy rozsądnym FDP.

	Bbetahat	SEbetahat	SEphat	T	F	FDP
GLM	0.7671610	5343.4586312	47.536711	0.00	0.00	NaN
BIC	-0.0019538	0.3282849	7.872553	26.05	2.30	0.0784440
mBIC	-0.0015615	0.2779893	5.384392	17.70	0.60	0.0313877
mBIC2	-0.0015803	0.3121813	7.043558	22.90	1.75	0.0693896
LASSO	-0.0034520	0.5955909	23.894404	31.30	37.60	0.5443732
SLOPE	-0.0000582	1.6734693	42.828540	47.85	131.00	0.7316004

W momencie w którym dodajemy korelację do modelu z 50-ma istotnymi zmiennymi sygnał jest jeszcze trudniejszy do odtworzenia. Funkcja glm ma astronomiczne wartości błędów i nie znajduje żadnych odkryć. Najwięcej prawdziwych odkryć wyłapuje SLOPE kosztem błędów estymacji i FDP. LASSO znajduje mniej odkryć, ale ma mniejsze błędy i mniejsze FDP od SLOPE. Najlepiej znowu sprawują się kryteria informacyjne, nieznacznie na plus wyróżnia się BIC za najwięcej prawdziwych odkryć przy małym FDP.

Zadanie 4

Podpunkt a

```
## [1] 0.9854646696 1.2391321200 0.7772665072 0.6739593671 0.2012132559
## [6] 0.4621502864 -0.2203136904 -0.1637088863 0.0003158717 -0.0147058657
```

```
## [11] 0.0711805195 -0.3419851141 -0.0075761734 -0.7066544020 0.4012132036
## [16] -0.5513103403 0.3086047194 0.3601994773 0.3147502423 0.6008827593
## [21] 0.1336718150 -0.7317522953 0.2016330217 -0.0980429430 -0.6451230726

## [1] "cv.glmnet"

## [1] 0.02568011 0.08446508 0.00000000 0.04862484 0.00000000
## [6] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
## [11] 0.00000000 0.00000000 0.00000000 -0.03922348 0.00000000
## [16] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
## [21] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000

## [1] "Regresja Firtha"

## [1] 0.628346504 0.798411380 0.515486819 0.427162192 0.148944865
## [6] 0.306350501 -0.124010106 -0.123145965 -0.003939801 -0.009868044
## [11] 0.077446538 -0.222049262 -0.018248232 -0.472645525 0.272394853
## [16] -0.341734092 0.206757328 0.207586805 0.218666121 0.382498536
## [21] 0.069943545 -0.473894955 0.121027731 -0.045529314 -0.379702840
```

Podpunkt b

	Bbetahat	SEbetahat	SEphat	Prawdziwe odkrycia	Falszywe odkrycia	FDP
GLM	-0.005	3.757	6.032	2	1	0.333
cv.glmnet	-0.003	0.228	10.388	3	1	0.250
Firth	-0.001	0.134	14.306	2	0	0.000

Podpunkt c

```
## [1] -0.096729369 0.353597717 0.208093541 0.004530668
```

	Bbetahat	SEbetahat	SEphat	Prawdziwe odkrycia	Falszywe odkrycia	FDP
mBIC2	-0.003	0.234	14.532	2	1	0.333
SLOPE	-0.003	0.234	14.532	2	1	0.333

Podpunkt d

	Bbetahat	SEbetahat	SEphat	T	F	FDP
GLM	2.8692773	1.099721e+04	6.533165	4.10	2.20	NaN
Firth	-0.0157312	1.767934e-01	5.184867	4.05	0.40	0.0751190
mBIC2	-0.0187207	1.667751e-01	8.815449	3.20	0.70	0.1639286
LASSO	-0.0111710	2.034286e-01	4.312483	4.50	2.95	0.3041828
SLOPE	-0.0183921	1.525888e-01	10.691484	2.70	0.25	NaN

Funkcja glm ma największe błędy estymacji z wyjątkiem SEphat Najmniejsze FDP ma regresja Firtha. LASSO i funkcja glm wykrywają najwięcej sygnału kosztem FDP.

Zadanie 5

Podpunkt a

```
## [1] "glm"

## [1] 675.157 705.072 255.639 304.405 268.870 279.190 -150.695
## [8] 439.416 117.900 -54.165 -130.884 -38.839 67.429 -26.660
## [15] -119.775 -240.684 -138.228 320.927 131.613 135.810 55.152
## [22] -356.823 114.820 -290.548 -420.559

## [1] "Regresja Firtha"

## [1] 1.31797032 0.97909679 0.34507640 1.04829691 0.84526488
## [6] 0.36424591 -0.56241538 0.86497853 -0.01820946 -0.36157266
## [11] 0.24994198 -0.16947401 0.66918772 0.02620209 -0.17827511
## [16] -0.16536090 -0.55576719 0.74559006 0.55156823 -0.26689573
## [21] -0.05521737 -0.73609213 0.25056346 -0.19517356 -0.80857352

## [1] "cv.glmnet"

## [1] 0.045 0.040 0.032 0.122 0.076 0.000 0.000 0.039 0.000 0.000 0.000
## [12] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [23] 0.000 0.000 0.000
```

Podpunkt b

	Bbetahat	SEbetahat	SEphat	Prawdziwe odkrycia	Falszywe odkrycia	FDP
GLM	75.962	2148747.181	11.380	0	0	NaN
cv.glmnet	-0.006	0.196	1.312	5	1	0.167
Firth	-0.021	0.170	13.734	2	0	0.000

Podpunkt c

```
## $est.beta
## [1] -0.07491317 0.25810377 0.37630717 0.24407456 0.22886979 0.23068425

## [1] 1 2 3 4 5 8 11
```

	Bbetahat	SEbetahat	SEphat	Prawdziwe odkrycia	Falszywe odkrycia	FDP
mBIC2	-0.006	0.196	1.312	5	0	0.000
SLOPE	-0.006	0.196	1.336	5	2	0.286

Podpunkt d

	Bbetahat	SEbetahat	SEphat	T	F	FDP
GLM	26.0968205	3.183451e+06	7.636652	2.05	1.15	NaN
Firth	-0.0241503	1.908485e-01	12.477645	2.30	0.60	0.1475000

	Bbetahat	SEbetahat	SEphat	T	F	FDP
mBIC2	-0.0155503	1.652547e-01	6.150728	3.95	1.20	0.1898016
LASSO	-0.0173825	1.796828e-01	5.952606	4.15	2.85	0.3458442
SLOPE	-0.0155720	2.115453e-01	5.547393	4.40	4.70	0.4483106

Funkcja glm na ponownie największe 2 błędy estymacji. Wprowadzenie korelacji utrudnia znalezienie sygnału. Funkcja glm i regresja Firtha znajdują poniżej połowy sygnału i przy niskim, ale niezadowalającym FDP. mBIC2 znajduje 80% sygnału przy podobnym FDP jak dwie poprzednie funkcje. LASSO i SLOPE znajdują najwięcej sygnału przy większe ilości fałszywych odkryć, LASSO ma mniejsze FDP.