

Regresja grzbietowa

Cel:

$$b_R = \operatorname{argmin}\left\{\sum_i (y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)^2 : \beta\right\},$$

pod warunkiem $\sum_{j=1}^p \beta_j^2 \leq c$

Jest to równoważne zagadnieniu

$$b_R = \operatorname{argmin}\left\{\sum_i (y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 : \beta\right\}.$$

Duże λ odpowiadają małym c

Reprezentacja macierzowa

$$b_R = (X'X + \lambda I)^{-1} X'y$$

Błąd średniokwadratowy estymatora ridge ma postać:

$$MSE(b_R) = \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + \lambda)^2} + \lambda^2 \beta'(X'X + \lambda I)\beta$$

Pierwszy składnik tego wzoru reprezentuje sumę wariancji b_R , drugi - sumę kwadratów obciążeń. Gdy λ rośnie pierwszy czynnik maleje, drugi rośnie. Celem jest znalezienie takiego λ by $MSE(b_R)$ miał wartości mniejsze niż wartość $MSE(b_R)$.

Parametr λ wyznacza się, obserwując wartości b_R dla ciągu wartości $0 \leq \lambda \leq 1$ i wybiera się taką wartość λ , dla której estymatory się stabilizują. Przyjmuje się, że dobrym oszacowaniem optymalnego λ są:

estymator k_{HKB} (estymator Hoerla i Kennarda):

$$k_{HKB} = \frac{ps^2}{b'b}$$

i estymator k_{LW} Lawlessa i Wanga:

$$k_{LW} = \frac{ps^2}{\sum_{j=1}^p \lambda_j b_j^2}$$

W powyższych wzorach:

λ_j są wartościami własnymi macierzy $X'X$,

s i b - estymatorami najmniejszych kwadratów dla σ i β .

PRZYKŁAD

```

> metabolizm.full\$coef
(Intercept)          g          F          A          g:F          g:A
-1.659660    2.514157    1.465719    2.552104   -1.673438   -1.458742
      F:A      g:F:A
-2.251711  1.198668
> vif(metabolizm.full)
          g          F          A          g:F          g:A          F:A          g:F:A
2.159685  8.889297 14.443701  6.148334 10.837870 33.363977 30.191160

> met.ridge <- lm.ridge(m~g*F*A, alkohol)
> select(met.ridge)
modified HKB estimator is 0.6940266
modified L-W estimator is 1.387960

> met.ridge0 <- lm.ridge(m~g*F*A,alkohol,lambda = 0.6940266)
> coef(met.ridge0)
          g          F          A          g:F          g:A          F:A          F:A
-1.0006138  2.2767315  0.5552916  1.2619662 -1.2824724 -0.8454159 -0.3667415
      g:F:A
0.1857708

> metabolizm.zero\$coef
(Intercept)          g          F          A
-0.2440174    1.9466239   -1.6535251   -0.1183464

> vif(metabolizm.zero)
          g          F          A
1.307574  1.306167  1.181513

> met0.ridge <-lm.ridge(m~g+F+A,alkohol)
> select(met0.ridge)
modified HKB estimator is 0.4484636
modified L-W estimator is 0.3496626

> met0.ridge0 <-lm.ridge(m~g+F+A,alkohol,lambda =0.3496626 )
> coef(met0.ridge0)
          g          F          A
-0.2016814  1.9248492 -1.6527207 -0.1272790

```

Metody oparte na współczynniku determinacji

Porównując modele regresji można posłużyć się dwoma współczynnikami:

- współczynnikiem determinacji

$$R^2 = \frac{d^2(M_p, M_1)}{d^2(D, M_1)} = 1 - \frac{d^2(D, M_p)}{d^2(D, M_1)}$$

- skorygowanym współczynnikiem determinacji

$$R_{adj}^2 = 1 - \frac{\frac{d^2(D, M_p)}{n-p}}{\frac{d^2(D, M_1)}{n-1}}$$

M_p i M_1 są modelami z odpowiednio p i 1 współczynnikiem (M_1 jest modelem stałym), D jest wektorem danych.

Współczynnik determinacji R^2 rośnie wraz z p , więc jest mało użytecznym wskaźnikiem dobrego modelu.

Skorygowany współczynnik determinacji R_{adj}^2 jest tym większy im błąd średniokwadratowy modelu M_p , równy $MSE(M_p) = \frac{d^2(D, M_p)}{n-p}$ jest mniejszy, co wskazuje na dobre dopasowanie modelu.

PRZYKŁAD

```
lm(m~g*F*A,data=alkohol)
```

Multiple R-squared: 0.8277, Adjusted R-squared: 0.7774

```
lm(m ~ g + F + g:F, data = alkohol)
```

Multiple R-squared: 0.8137, Adjusted R-squared: 0.7938

C_p Mallowsa

Ważony współczynnik błędu średniokwadratowego modelu M_p

$$\frac{MSE(M_p)}{\hat{\sigma}^2}$$

mierzy ile razy aktualny model jest lepszy od modelu pełnego ($\hat{\sigma}$ jest błędem średniokwadratowym modelu pełnego).

Interesujące są jedynie takie modele , że $MSE(M_p) \leq \hat{\sigma}^2$.

$$C_p = (n - p) \frac{MSE(M_p)}{\hat{\sigma}^2} - n + 2p = \frac{d^2(D, M_p)}{\hat{\sigma}^2} - n + 2p$$

C_p Mallowsa ma dwie ważne własności:

- $MSE(M_p) \leq \hat{\sigma}^2 \Leftrightarrow C_p \leq p$,
- Mała wartość C_p oznacza minimum $d^2(D, M_p)$ przy dodatkowym warunku $p = \min!$ (metoda funkcji kary)

Praktyczna zasada korzystania ze wskaźnika C_p Mallowsa:

- Wybrać najmniejsze C_p takie, że $C_p < p$,

PRZYKŁAD

```
> library("wle")
> metfull.cp <- mle.cp(metabolizm.full)
> summary(metfull.cp)
```

Call:

```
mle.cp(formula = metabolizm.full)
```

Mallows Cp:

	(Intercept)	g	F	A	g:F	g:A	F:A	g:F:A	cp
[1,]	0	1	0	0	1	0	0	0	0.5726
[2,]	1	1	0	0	1	0	0	0	0.7316
[3,]	0	1	0	0	1	1	0	0	1.3870
[4,]	0	1	0	1	1	0	0	0	1.9050
[5,]	1	1	1	0	1	0	0	0	1.9440
[6,]	1	1	0	0	1	1	0	0	2.3840
[7,]	0	1	1	0	1	0	0	0	2.5170
[8,]	0	1	0	0	1	0	0	1	2.5520
[9,]	0	1	0	0	1	0	1	0	2.5540
[10,]	1	1	0	1	1	0	0	0	2.7000
[11,]	1	1	0	0	1	0	1	0	2.7170
[12,]	1	1	0	0	1	0	0	1	2.7260
[13,]	0	1	0	1	1	1	0	0	2.9680
[14,]	0	1	0	0	1	1	0	1	3.0910
[15,]	0	1	0	0	1	1	1	0	3.1010
[16,]	1	1	0	1	1	1	0	0	3.2510
[17,]	0	1	1	0	1	1	0	0	3.3790
[18,]	0	1	0	1	1	0	1	0	3.6830
[19,]	0	1	0	1	1	0	0	1	3.7100
[20,]	1	1	1	0	1	1	0	0	3.7670

Printed the first 20 best models