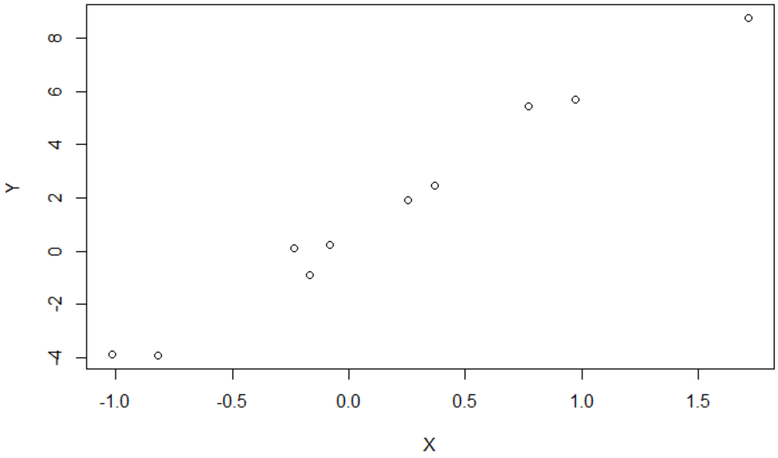
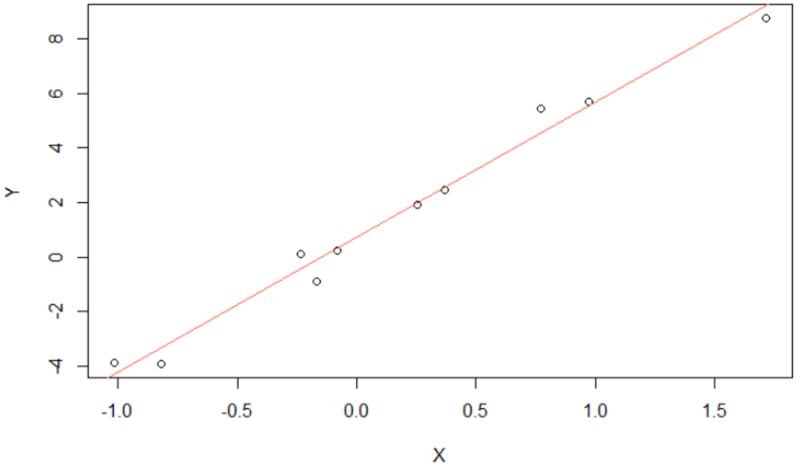
Wprowadzenie do modeli liniowych

Krystyna Grzesiak

Y	X
-3.89	-1.02
0.25	-0.08
0.13	-0.23
-3.90	-0.82
5.45	0.77
-0.88	-0.17
5.70	0.97
8.74	1.72
1.90	0.26
2.47	0.37



Y	X
-3.89	-1.02
0.25	-0.08
0.13	-0.23
-3.90	-0.82
5.45	0.77
-0.88	-0.17
5.70	0.97
8.74	1.72
1.90	0.26
2.47	0.37



$$y = b_1 x + b_0$$

Model liniowy Gaussa-Markowa

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \epsilon \tag{1}$$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Parametry:

- Y - wektor losowy o znanej realizacji (zmienna objaśniana)
- $\mathrm{X} = (1, X_1, \ldots, X_p)$ - macierz planu; $X_1, \ldots, X_p$ - znane wektory deterministyczne (zmienne objaśniajace)
- $\beta_0, \beta_1, \ldots, \beta_p$ - nieznany wektor deterministyczny (wektor współczynników modelu)
- ϵ - wektor losowy o nieznanej realizacji (szum)

Założenia

- liniowość - wzór 1
- $\epsilon_i \sim N(0, \sigma^2)$ (wariancja szumu jest stała dla każdego X_i)
- $Y \sim N(X\beta, \sigma^2)$
- $X_1, \ldots, X_p$ sa niezależne; macierz X jest pełnego wymiaru

Metoda najmniejszych kwadratów - estymacja wektora β

Definiujemy funkcje:

$$S(\beta) = S(\beta_0, \ldots, \beta_p) = ||Y - \mathrm{X}\beta||^2 = \sum_{i=1}^{n} \Big(y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}) \Big)^2,$$

wtedy

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{argmin} S(\beta)$$

oraz

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_p X_p.$$

Gdy X jest pełnego rzedu otrzymujemy jednoznaczne rozwiazanie

$$\hat{\beta} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'Y$$

i wówczas: $\mathrm{E}\hat{\beta} = \beta$ oraz $cov(\hat{\beta}) = \sigma^2(\mathrm{X}'\mathrm{X})^{-1}$.

Interpretacja geometryczna

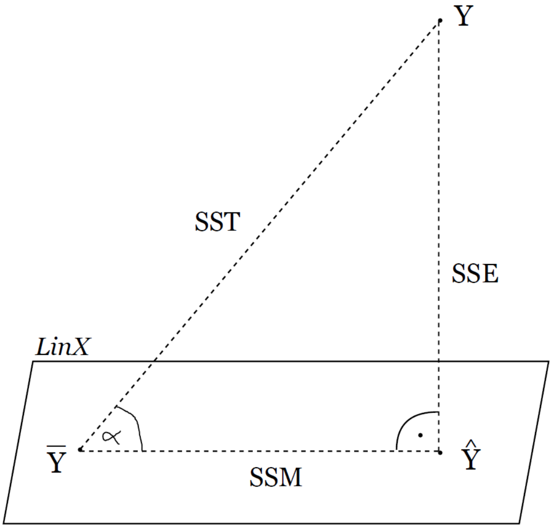
$$SST = ||Y - \overline{Y}||, dfT = n - 1$$

$$SSE = ||Y - \hat{Y}||, dfE = n - p$$

$$SSM = ||\overline{Y} - \hat{Y}||, dfM = p - 1$$

$$R^2 = \cos^2 \alpha$$

$$\hat{\sigma}^2 = \frac{SSE}{dfE}$$



Przykład

$$Y = \beta_0 + \beta_1 X + \epsilon$$

t-test:

Testujemy $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$

Statystyka testowa: $t = \dfrac{\hat{\beta}_1}{s(\beta_1)} \sim t_{n-2}$

F-test:

Testujemy $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$

Statystyka testowa: $F = \dfrac{SSM}{\frac{SSE}{n-2}} \sim F_{1,n-2}$

```
> model = lm(Y~X)
> summary(model)

Call:
lm(formula = Y ~ X)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.77887 -0.38738 -0.07302  0.35751  0.90479 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.7171     0.1822   3.936  0.00432 ** 
X             4.9602     0.2264  21.909 1.99e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5619 on 8 degrees of freedom
Multiple R-squared:  0.9836,	Adjusted R-squared:  0.9816 
F-statistic:   480 on 1 and 8 DF,  p-value: 1.988e-08
```

- R^2: Lepszy model $>$ Gorszy model
- $\hat{\sigma}^2$: Lepszy model $<$ Gorszy model
- F: Lepszy model $>$ Gorszy model

Porównanie modeli zagnieżdżonych

$$F = \frac{\frac{SSE(R) - SSE(F)}{dfE(R) - dfE(F)}}{\frac{SSE(F)}{dfE(F)}}.$$

Przy prawdziwości H_0 statystyka F ma rozkład Fishera–Snedecora F_{δ_1,δ_2} gdzie

$$\delta_1 = dfE(R) - dfE(F) = p_1 - p_2$$

jest liczba testowanych współczynników oraz

$$\delta_2 = dfE(F) = n - p_1.$$

Odrzucamy H_0 na poziomie istotności α kiedy

$$F > F^{-1}_{\delta_1,\delta_2}(1 - \alpha)$$

gdzie $F^{-1}_{\delta_1,\delta_2}(1 - \alpha)$ oznacza kwantyl rzedu $1 - \alpha$ z rozkładu F_{δ_1,δ_2}.

Kryteria informacyjne

Minimalizujemy:

- $\dfrac{RSS}{\sigma^2} + \pi(k, n, p)$ - znane σ^2
- $n \log(RSS) + \pi(k, n, p)$ - nieznane σ^2
- $-2 \log L + \pi(k, n, p)$ - uogólnione modele

"Kara na wymiar" dla poszczególnych kryteriów

- AIC: $2k$
- BIC: $k \log n$
- RIC: $2k \log p$
- mBIC: $k(\log n + 2 \log(\frac{p}{c}))$

Regularyzacja

$$Ridge: \underset{b\in\mathbb{R}^p}{argmin}\frac{1}{2}S(b) + \sum_{i=1}^{p}\lambda_i b^2$$

$$LASSO: \underset{b\in\mathbb{R}^p}{argmin}\frac{1}{2}S(b) + \lambda\sum_{i=1}^{p}|b_i|$$

$$SLOPE: \underset{b\in\mathbb{R}^p}{argmin}\frac{1}{2}S(b) + \sum_{i=1}^{p}\lambda_i|b|_{(i)}$$

gdzie $b_{(i)}$ oznacza i-ta statystyke pozycyjna oraz $0 \leq \lambda_p \leq \ldots \leq \lambda_2 \leq \lambda_1$.

Normy L1 i L2

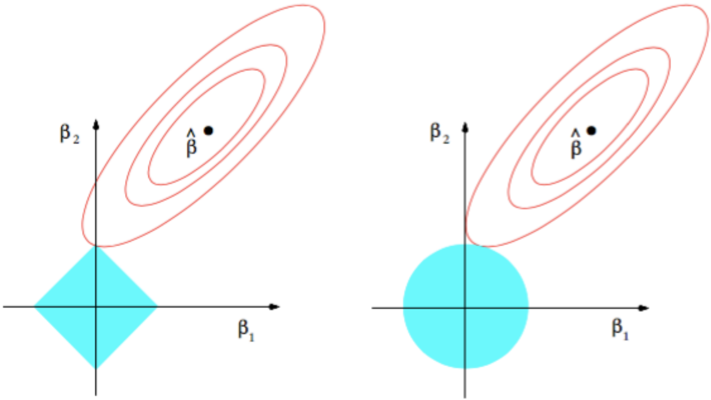


FIGURE 3.11. *Estimation picture for the lasso (left) and ridge regression (right). Shown are contours of the error and constraint functions. The solid blue areas are the constraint regions* $|\beta_1| + |\beta_2| \leq t$ *and* $\beta_1^2 + \beta_2^2 \leq t^2$*, respectively, while the red ellipses are the contours of the least squares error function.*

Dziekuje za uwage