O estimador Ridge Regression

$$\tilde{\beta} = \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta)$$

sujeito a: $\sum_{i=1}^{p} \beta_i^2 \leq c$

O estimador LASSO

Least Absolute Shrinkage Selection Operator

$$\tilde{\beta} = \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta)$$

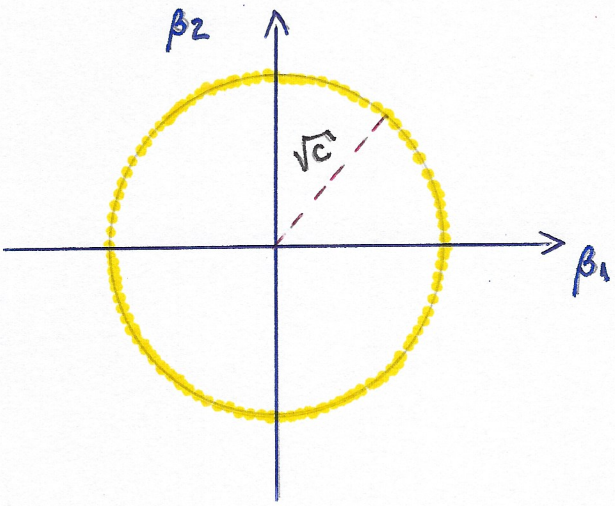
sujeito a: $\sum_{i=1}^{p} |\beta_i| \leq c$

Ridge ou LASSO — $\sum_{i=1}^{p} \beta_i^2$ ou $\sum_{i=1}^{p} |\beta_i|$?

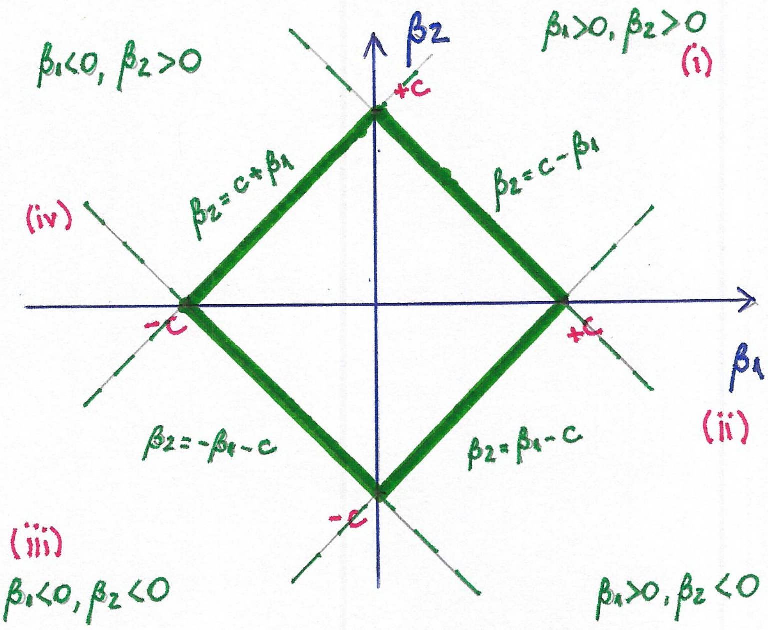
vamos supor $p = 2$ → $\sum \beta_i^2 = \beta_1^2 + \beta_2^2 \leq c$ (a)

→ $\sum |\beta_i| = |\beta_1| + |\beta_2| \leq c$ (b)

(a) $\beta_1^2 + \beta_2^2 = c$



(b) $|\beta_1| + |\beta_2| = c$



(i) $\beta_1 > 0$ e $\beta_2 > 0$ → $|\beta_1| + |\beta_2| = c$ → $\beta_1 + \beta_2 = c$ → $\beta_2 = c - \beta_1$

(ii) $\beta_1 > 0$ e $\beta_2 < 0$ → $|\beta_1| + |\beta_2| = c$ → $\beta_1 - \beta_2 = c$ → $\beta_2 = \beta_1 - c$

(iii) $\beta_1 < 0$ e $\beta_2 < 0$ → $|\beta_1| + |\beta_2| = c$ → $-\beta_1 - \beta_2 = c$ → $\beta_2 = -\beta_1 - c$

(iv) $\beta_1 < 0$ e $\beta_2 > 0$ → $|\beta_1| + |\beta_2| = c$ → $-\beta_1 + \beta_2 = c$ → $\beta_2 = c + \beta_1$

então:

$\sum_{i=1}^{p} |\beta_i| \leq c$ pode ser escrito como um conjunto de restrições lineares na forma $C\beta \leq D$

no exemplo: $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$

$$\underbrace{\begin{bmatrix} +1 & +1 \\ +1 & -1 \\ -1 & -1 \\ -1 & +1 \end{bmatrix}}_{C} \times \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}}_{\beta} \leq \underbrace{\begin{bmatrix} c \\ c \\ c \\ c \end{bmatrix}}_{D}$$

pg. 92:

$$\tilde{\beta} = \hat{\beta} - (X^T X)^{-1} C^T \left[C (X^T X)^{-1} C^T \right]^{-1} \left[C\hat{\beta} - D \right]$$

$$H = X (X^T X)^{-1} \left\{ X^T - C^T \left[C (X^T X)^{-1} C^T \right]^{-1} C (X^T X)^{-1} X^T \right\}$$

lembrando que: $PRESS = \sum_{i=1}^{n} \left(\frac{r_i}{1-h_{ii}} \right)^2$

$$R^2_{Pred} = 1 - \frac{PRESS}{SQ_T}$$

neste caso, o valor do R^2_{Pred} é sensível ao parâmetro c (minúsculo).