The Elements of statistical learning

Ex. 3.5 Consider the ridge regression problem (3.41). Show that this problem is equivalent to the problem

$$\hat{\beta}^c = \underset{\beta^c}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \left[y_i - \beta_0^c - \sum_{j=1}^{p} (x_{ij} - \bar{x}_j)\beta_j^c \right]^2 + \lambda \sum_{j=1}^{p} {\beta_j^c}^2 \right\}. \quad (3.85)$$

Give the correspondence between β^c and the original β in (3.41). Characterize the solution to this modified criterion. Show that a similar result holds for the lasso.

Remplaçons $x_{ij} \Rightarrow x_{ij} - \bar{x}_j$ alors :

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \left[y_i - \beta_0 - \sum_{j=1}^{p} \bar{x}_j \beta_p - \sum_{j=1}^{p} (x_{ij} - \bar{x}_j)\beta_j \right]^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}.$$

Or :

$$\beta_0^c = \beta_0 + \sum_{j=1}^{p} \bar{x}_j \beta_j$$

$$\beta_j^c = \beta_j \text{ pour } j = 1, \dots, p$$

Donc on peut voir que la même technique de centrage s'applique sur les 2.

Maintenant "caractérisons" la solution :

dérivons β_0^c

$$\sum_{i=1}^{N}\left(y_i - \beta_0^c - \underbrace{\sum_{j=1}^{p}(x_{ij} - \bar{x}_j)\beta_j}_{=0}\right) = 0$$

Donc on peut le simplifier par :

$$\sum_{i=1}^{N}(y_i - \beta_0^c)$$

$$\Leftrightarrow \sum y_i - n\beta_0^c = 0$$

$$\beta_0^c = \frac{\sum y_i}{n} = \bar{y}$$

donc $\beta_0^c = \bar{y}$

$$\tilde{y}_i = y_i - \beta_0^c$$

$$\tilde{x}_{ij} = x_{ij} - \bar{x}_j$$

$$\min_{\beta_c} (\tilde{y} - \tilde{X}\beta^c)^T(\tilde{y} - \tilde{X}\beta^c) + \lambda \beta_c^T \beta_c$$

$$\hat{\beta}_c = (\tilde{X}^T\tilde{X} + \lambda I)^{-1}\tilde{X}^T y.$$