minimizing over ξ, w and $\widehat{b}$. If the solution to this program is $\widehat{w}$, then $\widehat{b}$ is given by

$$\widehat{b} = \overline{\widehat{y}} - (\overline{\widehat{X}^1} \cdots \overline{\widehat{X}^n})\widehat{w} = 0,$$

since the data $\widehat{y}$ and $\widehat{X}$ are centered. Therefore (**RR6**) *is equivalent to ridge regression without an intercept term applied to the centered data* $\widehat{y} = y - \overline{y}\mathbf{1}$ *and* $\widehat{X} = X - \overline{X}$,

Program (RR6′):

$$\begin{aligned} &\text{minimize} \quad \xi^\top \xi + K w^\top w \\ &\text{subject to} \\ &\qquad\qquad \widehat{y} - \widehat{X} w = \xi, \end{aligned}$$

minimizing over ξ and w.

If $\widehat{w}$ is the optimal solution of this program given by

$$\widehat{w} = \widehat{X}^\top (\widehat{X}\widehat{X}^\top + K I_m)^{-1} \widehat{y}, \qquad (*_{w6})$$

then b is given by

$$b = \overline{y} - (\overline{X^1} \cdots \overline{X^n})\widehat{w}.$$

Remark: Although this is not obvious a priori, the optimal solution w^* of the Program (**RR3**) given by $(*_{w_3})$ is equal to the optimal solution $\widehat{w}$ of Program (**RR6′**) given by $(*_{w_6})$. We believe that it should be possible to prove the equivalence of these formulae but a proof eludes us at this time. We leave this as an open problem. In practice the Program (**RR6′**) involving the centered data appears to be the preferred one.

Example 55.1. Consider the data set (X, y_1) with

$$X = \begin{pmatrix} -10 & 11 \\ -6 & 5 \\ -2 & 4 \\ 0 & 0 \\ 1 & 2 \\ 2 & -5 \\ 6 & -4 \\ 10 & -6 \end{pmatrix}, \quad y_1 = \begin{pmatrix} 0 \\ -2.5 \\ 0.5 \\ -2 \\ 2.5 \\ -4.2 \\ 1 \\ 4 \end{pmatrix}$$

as illustrated in Figure 55.1. We find that $\overline{y} = -0.0875$ and $(\overline{X^1}, \overline{X^2}) = (0.125, 0.875)$. For the value $K = 5$, we obtain

$$w = \begin{pmatrix} 0.9207 \\ 0.8677 \end{pmatrix}, \quad b = -0.9618,$$