where $\overline{y}$ is the mean of y and $\overline{X^j}$ is the mean of the jth column of X. Therefore,

$$b = \overline{y} - \sum_{j=1}^{n} \overline{X^j} w_j = \overline{y} - (\overline{X^1} \ \cdots \ \overline{X^n}) w,$$

where $(\overline{X^1} \ \cdots \ \overline{X^n})$ is the $1 \times n$ row vector whose jth entry is $\overline{X^j}$.

We will now show that solving the dual (**DRR3**) for α and obtaining $w = X^\top \alpha$ is equivalent to solving our previous ridge regression Problem (**RR2**) applied to the centered data $\widehat{y} = y - \overline{y}\mathbf{1}_m$ and $\widehat{X} = X - \overline{X}$, where $\overline{X}$ is the $m \times n$ matrix whose jth column is $\overline{X^j}\mathbf{1}_m$, the vector whose coordinates are all equal to the mean $\overline{X^j}$ of the jth column X^j of X.

The expression

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

suggests looking for an intercept term b (also called bias) of the above form, namely

Program (RR4):

$$\begin{aligned} &\text{minimize} \quad \xi^\top \xi + K w^\top w \\ &\text{subject to} \\ &\qquad y - Xw - b\mathbf{1} = \xi \\ &\qquad b = \widehat{b} + \overline{y} - (\overline{X^1} \ \cdots \ \overline{X^n}) w, \end{aligned}$$

with $\widehat{b} \in \mathbb{R}$. Again, in Program (**RR4**), minimization is performed over ξ, w, b and $\widehat{b}$, but b and $\widehat{b}$ are not penalized.

Since

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

if $\overline{X} = (\overline{X^1}\mathbf{1} \ \cdots \ \overline{X^n}\mathbf{1})$ is the $m \times n$ matrix whose jth column is the vector $\overline{X^j}\mathbf{1}$, then the above program is equivalent to the program

Program (RR5):

$$\begin{aligned} &\text{minimize} \quad \xi^\top \xi + K w^\top w \\ &\text{subject to} \\ &\qquad y - Xw - \overline{y}\mathbf{1} + \overline{X} w - \widehat{b}\mathbf{1} = \xi, \end{aligned}$$

where minimization is performed over ξ, w and $\widehat{b}$. If we write $\widehat{y} = y - \overline{y}\mathbf{1}$ and $\widehat{X} = X - \overline{X}$, then the above program becomes

Program (RR6):

$$\begin{aligned} &\text{minimize} \quad \xi^\top \xi + K w^\top w \\ &\text{subject to} \\ &\qquad \widehat{y} - \widehat{X} w - \widehat{b}\mathbf{1} = \xi, \end{aligned}$$