Program (Dual Elastic Net):

$$\begin{aligned}&\text{minimize} \quad \frac{1}{2}\begin{pmatrix}\beta_+^\top & \beta_-^\top & \mu_+^\top & \mu_-^\top\end{pmatrix} P \begin{pmatrix}\beta_+\\ \beta_-\\ \mu_+\\ \mu_-\end{pmatrix} + q^\top \begin{pmatrix}\beta_+\\ \beta_-\\ \mu_+\\ \mu_-\end{pmatrix}\\ &\text{subject to}\\ &\qquad A\begin{pmatrix}\beta_+\\ \beta_-\\ \mu_+\\ \mu_-\end{pmatrix} = c,\end{aligned}$$

$\beta_+, \beta_- \in \mathbb{R}^n_+, \mu_+, \mu_- \in \mathbb{R}^m_+$.

Once $\xi = K\mu = K(\mu_+ - \mu_-)$ and w are determined by $(*_w)$, we obtain b using the equation

$$b\mathbf{1}_m = y - Xw - \xi,$$

which as in Section 55.5 yields

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

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

We leave it as an easy exercise to show that A has rank $n+1$. Then we can solve the above program using ADMM, and we have done so. This very similar to what is done in Section 56.3, and hence the details are left as an exercise.

Observe that when $\tau = 0$, the elastic net method reduces to ridge regression. As K tends to 0 the elastic net method tends to lasso, but $K = 0$ is not an allowable value since $\tau/0$ is undefined. Anyway, if we get rid of the constraint

$$\beta_+ + \beta_- = \frac{\tau}{K}\mathbf{1}_n$$

the corresponding optimization program not does determine w. Experimenting with our program we found that convergence becomes very slow for $K < 10^{-3}$. What we can do if K is small, say $K < 10^{-3}$, is to run lasso. A nice way to "blend" ridge regression and lasso is to call the elastic net method with $K = C(1-\theta)$ and $\tau = C\theta$, where $0 \leq \theta < 1$ and $C > 0$.

Running the elastic net method on the data set $(X14, y14)$ of Section 55.5 with $K = \tau = 0.5$ shows absolutely no difference, but the reader should conduct more experiments to see how elastic net behaves as K and τ are varied (the best way to do this is to use θ as explained above). Here is an example involving a data set $(X20, y20)$ where $X20$ is a 200×30 matrix generated as follows: