Regularized Regression

David Fleischer

Last Update 2017-12-04

Introduction

We consider data (X, Y), $X \in \mathbb{R}^{n \times p}$, $Y \in \mathbb{R}^n$ such that the responses Y are linearly related to predictors X through coefficients $\beta \in \mathbb{R}^p$, i.e.

$$Y = X\beta + \epsilon$$

where $\epsilon \in \mathbb{R}^n$ is some (unobservable) random perturbation with zero mean, constant variance, and zero covariance

$$\begin{aligned}
\mathbb{E}\left[\epsilon\right] &= \mathbf{0} \in \mathbb{R}^n \\
\mathrm{Var}\left(\epsilon\right) &= \sigma^2 \mathbb{I}_p \in \mathbb{R}^{p \times p}, \quad \sigma^2 \in \mathbb{R}
\end{aligned}$$

We wish to investigate problem of estimating the coefficients β by the *regularized/penalized* regression estimates $\hat\beta^{(q)}$ given by

$$\hat\beta^{(q)} = \arg\min_\beta \left\{ \frac{1}{2} \|Y - X\beta\|_2^2 \right\}, \quad \|\beta\|_q^q \leq t^q, \quad t > 0 \tag{1}$$

where $\|x\|_q^q$, $x \in \mathbb{R}^r$, $q > 0$, is given by

$$\|x\|_q^q = \sum_{j=1}^r |x_j|^q$$

For $q > 1$ this quantity is the ℓ_q norm, while for $q \in (0, 1)$ this quantity is *not* considered a norm as a consequence of failing the triangle inequality. We simplify our objective function in (1) by

$$\begin{aligned}
\arg\min_\beta \left\{ \frac{1}{2} \|Y - X\beta\|_2^2 \right\} &= \arg\min_\beta \left\{ \frac{1}{2} \|Y - X\beta\|_2^2 \right\} \\
&\equiv \arg\min_\beta \left\{ \frac{1}{2} \left(Y - X\beta\right)^T \left(Y - X\beta\right) \right\} \\
&= \arg\min_\beta \left\{ \frac{1}{2} Y^T Y - \beta^T X^T Y + \frac{1}{2} \beta^T X^T X \beta \right\} \\
&= \arg\min_\beta \left\{ -\beta^T X^T Y + \frac{1}{2} \beta^T X^T X \beta \right\}
\end{aligned}$$

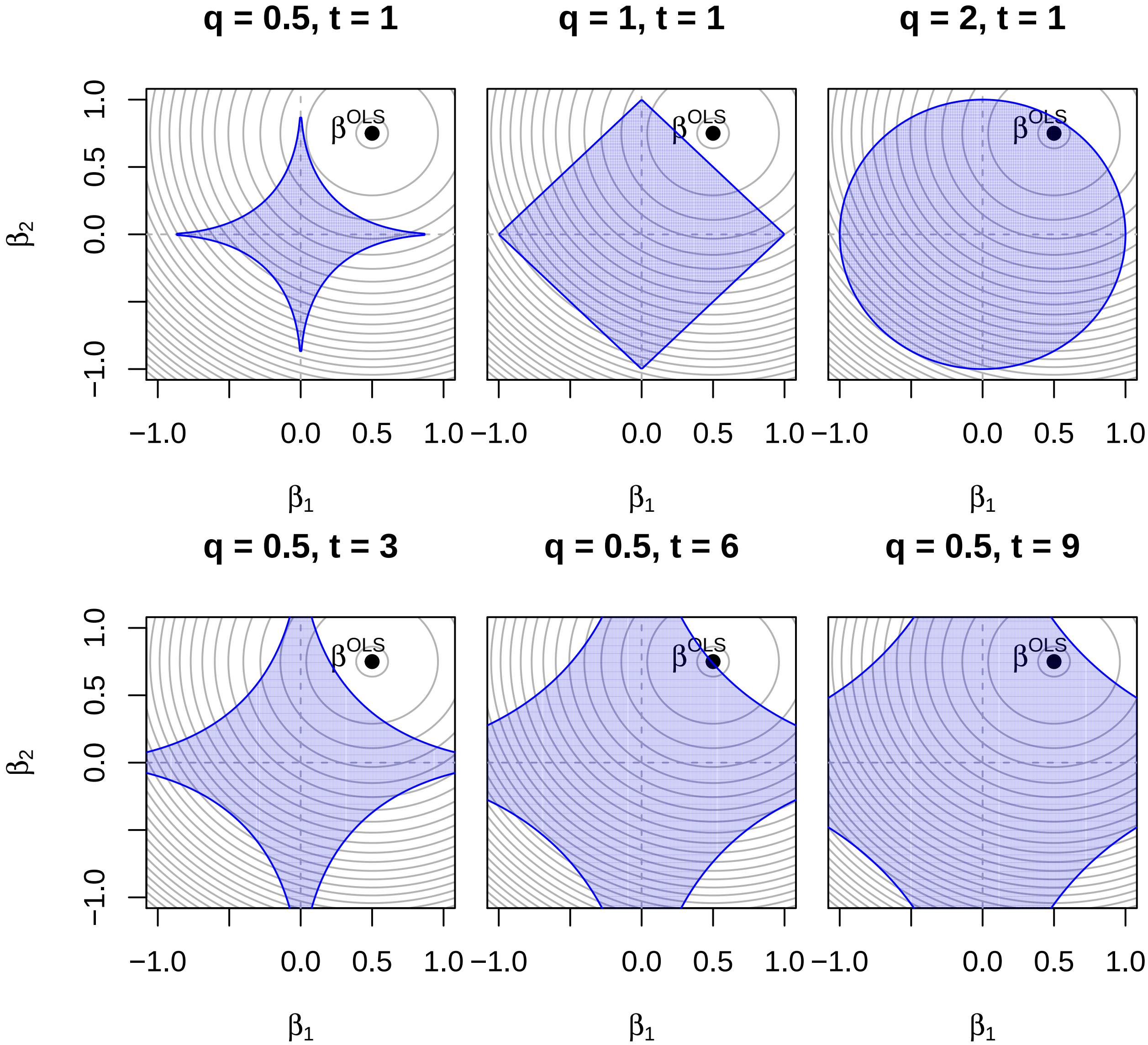
As an initial assumption we suppose our predictors are orthogonal $X^T X = \mathbb{I}_p$, yielding $\hat\beta^{\mathrm{OLS}} = X^T Y$. Therefore, we may yet again simplify (1) by

$$\arg\min_{\beta}\left\{\frac{1}{2}\|Y-X\beta\|_2^2\right\} = \arg\min_{\beta}\left\{-\beta^T X^T Y + \frac{1}{2}\beta^T X^T X\beta\right\}$$
$$= \arg\min_{\beta}\left\{-\beta^T\beta^{\mathrm{OLS}} + \frac{1}{2}\beta^T\beta\right\}$$

and so we restate our constrained optimization problem (1) as

$$\hat{\beta}^{(q)} = \arg\min_{\beta}\left\{-\beta^T\beta^{\mathrm{OLS}} + \frac{1}{2}\beta^T\beta\right\}, \quad \|\beta\|_q^q \le t^q, \quad t > 0 \tag{2}$$

Our objective function in (2) defines concentric ellipsoids in $\mathbb{R}^p$ centered at $\hat{\beta}^{\mathrm{OLS}}$ while our constraint defines "norms"[1] in $\mathbb{R}^p$. For example, when $p = 2$ we may easily visualize our optimization problem as finding the values of $\hat{\beta}$ closest to the center of the ellipsoids $\hat{\beta}^{\mathrm{OLS}}$ such that the pair $(\hat{\beta}_1, \hat{\beta}_2)$ is in the region defined by the q-penalty:



[1] If we permit the perversion of a norm to include the cases of $q \in (0, 1)$.

We can visualize the paths that our estimates $\hat{\beta}^{(q)}$ take as t varies