Elastic net projections

John Duchi

December 23, 2009

Consider the following elastic net (Zou and Hastie, 2005) projection problem:

$$\underset{\boldsymbol{x}}{\text{minimize}} \ \frac{1}{2}\left\|\boldsymbol{x}-\boldsymbol{v}\right\|_2^2 \quad \text{subject to} \quad \left\|\boldsymbol{x}\right\|_1 + \frac{\lambda}{2}\left\|\boldsymbol{x}\right\|_2^2 \le c \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^d$. Due to the symmetry in the objective and constraints, we can without loss of generality assume that $\boldsymbol{v} \succeq \boldsymbol{0}$, that is, $v_i \ge 0$ for each v_i, and flip signs of the resulting solution at the end. That is, we want to solve

$$\underset{\boldsymbol{x}}{\text{minimize}} \ \frac{1}{2}\left\|\boldsymbol{x}-\boldsymbol{v}\right\|_2^2 \quad \text{subject to} \quad \boldsymbol{1}^T\boldsymbol{x} + \frac{\lambda}{2}\left\|\boldsymbol{x}\right\|_2^2 \le c, \ \boldsymbol{x} \succeq \boldsymbol{0}. \tag{2}$$

In the above, we assume that $\boldsymbol{v} \succeq \boldsymbol{0}$ and that $\boldsymbol{1}^T\boldsymbol{v} + \frac{\lambda}{2}\left\|\boldsymbol{v}\right\|_2^2 > c$, otherwise the problem is solved by setting $\boldsymbol{x} = \boldsymbol{v}$. Introduce Lagrange multipliers $\theta \ge 0$ for the first inequality constraint and $\boldsymbol{\nu} \in \mathbb{R}^d$, $\boldsymbol{\nu} \succeq 0$ for the second. This gives us Lagrangian

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\nu}, \theta) = \frac{1}{2}\left\|\boldsymbol{x}-\boldsymbol{v}\right\|_2^2 + \theta\left(\boldsymbol{1}^T\boldsymbol{x} + \frac{\lambda}{2}\left\|\boldsymbol{x}\right\|_2^2 - c\right) + \boldsymbol{\nu}^T\boldsymbol{x}.$$

To solve the problem, we find the minimizing $\boldsymbol{x}$ of the Lagrangian by taking the gradient of $\mathcal{L}$ with respect to $\boldsymbol{x}$. This gives us

$$\nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}, \boldsymbol{\nu}, \theta) = \boldsymbol{x} - \boldsymbol{v} + \theta\boldsymbol{1} + \theta\lambda\boldsymbol{x} + \boldsymbol{\nu}.$$

By setting the above to $\boldsymbol{0}$ and solving for $\boldsymbol{x}$, we see that as a function of $\theta, \boldsymbol{\nu}$, we have the optimal $\boldsymbol{x}$ is equal to

$$\boldsymbol{x}(\theta, \boldsymbol{\nu}) = \left(\frac{1}{1+\theta\lambda}\right)(\boldsymbol{v} - \theta\boldsymbol{1} + \boldsymbol{\nu}).$$

Note that standard use of complimentarity conditions (Boyd and Vandenberghe, 2004) implies that $\nu_i > 0$ only when $x_i(\theta, \boldsymbol{\nu}) = 0$, so we have as in Duchi et al. (2008) that the solution $\boldsymbol{x}$ depends only on θ:

$$x_i(\theta) = \frac{\left[v_i - \theta\right]_+}{1+\theta\lambda}, \tag{3}$$

where $\left[\alpha\right]_+ = \max\{\alpha, 0\}$. That is, we are faced as in Duchi et al. with the following root finding problem: we must find the unique $\theta \ge 0$ such that

$$\sum_{i=1}^d \frac{\left[v_i - \theta\right]_+}{1+\theta\lambda} = c. \tag{4}$$

Note that the term on the left of Eq. (4) is a decreasing function of θ and so there is a unique root, which can be found using essentially identical techniques as in Duchi et al. (2008). Given the θ solving Eq. (4), we set $\boldsymbol{x}$ as in Eq. (3).

References

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the ℓ_1-ball for learning in high dimensions. In *Proceedings of the 25th International Conference on Machine Learning*, 2008.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, 67:301–320, 2005.