How can we project to the Elastic Net ball?

1 Major Lessons

In the last post we saw that Moreau decomposition may give us a way on projecting onto norm balls. But the problem is that the elastic net is not a norm (why?). So to project on it we need another method.

The way that John Duchi projects onto l_1-norm ball may be a good line of attack. Here are the steps that he uses to give a good algorithm for this:

1.1 Projecting onto l_1-norm Ball

-

1.1.1 Projecting onto a Simplex

1.1.2 Why it Works

1.2 Projecting onto Elastic Net Ball

Here is the projection problem:

$$\mathbf{w} = \underset{\mathbf{y}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{v} - \mathbf{y}\|_2^2, \quad \|\mathbf{y}\|_1 + \frac{1}{2}\|\mathbf{y}\|_2^2 \leq b.$$

Like before, from the symmetry of the elastic net ball, we can consider the following problem and compute the solution of the original problem from $w_i = \textbf{sgn}(v_i)u_i$:

$$\mathbf{u} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2}\||\mathbf{v}| - \mathbf{x}\|_2^2, \quad \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x}\|_2^2 \leq b, \mathbf{x} \succeq 0$$

where $(|\mathbf{v}|)_i = |v_i|$. Writing the Lagrangian:

$$\mathcal{L} = \frac{1}{2}\||\mathbf{v}| - \mathbf{x}\|_2^2 + \alpha(\mathbf{x}^T\mathbf{1} + \frac{1}{2}\|\mathbf{x}\|_2^2 - b) + \boldsymbol{\beta}^T\mathbf{x}$$

Taking derivative w.r.t. x_i and setting it to zero:

$$u_i = \frac{|v_i| - \alpha - \beta_i}{1 + \alpha}$$

If $u_i > 0$ complementary slackness tells us that $\beta_i = 0$, so:

$$u_i = \frac{|v_i| - \alpha}{1 + \alpha}$$

Assume that we know the number of non-zero elements of the optimal solution $\mathbf{u}$ to be ρ. Therefore $\forall i \in [\rho] : |v_{(i)}| > \alpha$ Then we can write the feasibility of $\mathbf{u}$ as:

$$\begin{aligned}
b &= \sum_{i=1}^{\rho} u_i + \frac{1}{2}\sum_{i=1}^{\rho} u_i^2 \\
&= \frac{1}{1+\alpha}\sum_{i=1}^{\rho}(|v_i| - \alpha) + \frac{1}{2(1+\alpha)}(|v_i| - \alpha)^2 \\
&= \frac{1}{2(1+\alpha)^2}\sum_{i=1}^{\rho}(2|v_i| + 2\alpha|v_i| - 2\alpha - 2\alpha^2) + (|v_i|^2 + \alpha^2 - 2\alpha|v_i|) \\
&= \frac{1}{2(1+\alpha)^2}\sum_{i=1}^{\rho}\left[|v_i|^2 + 2|v_i| + 1 - (\alpha^2 + 2\alpha + 1)\right] \\
&= \frac{1}{2(1+\alpha)^2}\sum_{i=1}^{\rho}(|v_i| + 1)^2 - \frac{\rho}{2}
\end{aligned}$$

Assuming ρ is known, we get the following for the α:

$$\alpha = \sqrt{\frac{1}{2b+\rho}\sum_{i=1}^{\rho}(|v_i|+1)^2} - 1$$

Now, we claim that ρ is the following:

$$\rho = \max\left\{ j \in [p] : v_{(j)} - \frac{1}{j}\left(\sum_{i=1}^{j}\left[v_{(i)} + \frac{v_{(i)}^2}{2}\right] - b\right) > 0 \right\}$$

Let's assume that we are wrong. Then the correct number of non-zero elements of the actual solution $\mathbf{u}^*$ is ρ^* with the corresponding Lagrange multiplier α^*. We know that still we have the following:

$$\forall i \leq \rho^* : u_{(i)}^* = \frac{|v_{(i)}| - \alpha^*}{1+\alpha^*} > 0, \quad \alpha^* = \sqrt{\frac{1}{2b+\rho^*}\sum_{i=1}^{\rho^*}(|v_{(i)}|+1)^2} - 1$$

First consider the case of $\rho < \rho^*$. Note that in such a case, $\alpha^* < \alpha$.