Alternating Direction Method of Multipliers

Lasso regression invokes the following optimization problem

$$\text{minimize} \quad \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

Rewrite this into ADMM form

$$\text{minimize} \quad \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\gamma\|_1$$

$$\text{subject to} \quad \beta - \gamma = 0$$

Thus, the augmented Lagrangian is

$$L_\rho(\beta,\gamma,u) = \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\gamma\|_1 + u^T(\beta-\gamma) + \frac{\rho}{2}\|\beta-\gamma\|_2^2$$

Which means that the Lasso ADMM consists of the following iterations:

$$\beta^{k+1} := \arg\min_\beta L_\rho(\beta,\gamma^k,u^k)$$

$$\gamma^{k+1} := \arg\min_\gamma L_\rho(\beta^{k+1},\gamma,u^k)$$

$$u^{k+1} := u^k + \rho(\beta^{k+1} - \gamma^{k+1})$$

To obtain the argmin's, we need to take the gradient of the objective function and set it equal to zero.

$$\nabla_\beta L_\rho(\beta,\gamma^k,u^k) = \nabla_\beta\{\frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\gamma^k\|_1 + (u^k)^T(\beta-\gamma^k) + \frac{\rho}{2}\|\beta-\gamma^k\|_2^2\}$$

$$= X^T(X\beta - y) + u^k + \rho(\beta - \gamma^k) \overset{set}{=} 0$$

$$\Rightarrow \beta^{k+1} = (X^TX + \rho I)^{-1}(X^Ty + \rho\gamma^k - u^k)$$

And

$$\gamma^{k+1} = \arg\min_\gamma L_\rho(\beta^{k+1},\gamma,u^k)$$

$$= \arg\min_\gamma\{\frac{1}{2}\|y - X\beta^{k+1}\|_2^2 + \lambda\|\gamma\|_1 + (u^k)^T(\beta^{k+1}-\gamma) + \frac{\rho}{2}\|\beta^{k+1}-\gamma\|_2^2\}$$

$$= \arg\min_\gamma\{\lambda\|\gamma\|_1 - (u^k)^T\gamma + \frac{\rho}{2}\|\beta^{k+1}-\gamma\|_2^2\}$$

$$= \arg\min_\gamma\{\lambda\|\gamma\|_1 - (u^k)^T\gamma + \frac{\rho}{2}\gamma^T\gamma - \rho(\beta^{k+1})^T\gamma\}$$

$$= \arg\min_\gamma\{\lambda\|\gamma\|_1 - (u^k + \rho\beta^{k+1})^T\gamma + \frac{\rho}{2}\gamma^T\gamma\}$$

$$= \arg\min_\gamma\{\frac{\lambda}{\rho}\|\gamma\|_1 - (\frac{1}{\rho}u^k + \beta^{k+1})^T\gamma + \frac{1}{2}\gamma^T\gamma\}$$

$$= \arg\min_\gamma\{\frac{\lambda}{\rho}\|\gamma\|_1 + \frac{1}{2}\|\gamma - (\frac{1}{\rho}u^k + \beta^{k+1})\|_2^2\}$$

$$= S_{\lambda/\rho}(\frac{1}{\rho}u^k + \beta^{k+1}) \quad \text{(see the proof of exercise06)}$$

Let $v_k = u_k / \rho$, we have

$$\gamma^{k+1} = S_{\lambda/\rho}(\beta^{k+1} + v^k) = \underset{\tau=1}{prox} \frac{\lambda}{\rho} \| \beta^{k+1} + v^k \|$$

We can also use scaled augmented Lagrangian to obtain the same solutions as in the paper (http://stanford.edu/~boyd/papers/pdf/admm_distr_stats.pdf). Finally, we have the following updates:

$$\beta^{k+1} := (X^T X + \rho I)^{-1}[X^T y + \rho(\gamma^k - v^k)]$$
$$\gamma^{k+1} := S_{\lambda/\rho}(\beta^{k+1} + v^k)$$
$$v^{k+1} := v^k + \beta^{k+1} - \gamma^{k+1}$$

Lastly, the stopping rules are

$$\| r^k \|_2 \le \varepsilon^{pri}, \qquad r^{k+1} = \beta^{k+1} - \gamma^{k+1} \qquad for\ \varepsilon^{pri} > 0$$
$$\| s^k \|_2 \le \varepsilon^{dual}, \qquad s^{k+1} = -\rho(\gamma^{k+1} - \gamma^k) \qquad for\ \varepsilon^{dual} > 0$$

Where ε^{pri} and ε^{dual} are feasibility tolerances for primal and dual feasibility conditions. These tolerances can be chosen using an absolute and relative criterion, such as

$$\varepsilon^{pri} = \sqrt{p}\varepsilon^{abs} + \varepsilon^{rel} \max\{\| \beta^k \|_2, \| \gamma^k \|_2\} \qquad where\ \varepsilon^{abs} > 0 \quad and\ \varepsilon^{rel} > 0$$
$$\varepsilon^{dual} = \sqrt{n}\varepsilon^{abs} + \varepsilon^{rel} \| v \|_2$$

Below are the ADMM lasso objective function plot vs. number of iterations with varying penalty parameters.