# 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.

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

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

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

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:

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

Lastly, the stopping rules are

$$\begin{aligned}
&\| r \|_2 < \varepsilon_r, \qquad r := \beta - \gamma \qquad for \;\; \varepsilon_r > 0 \\
&\| s \|_2 < \varepsilon_s, \qquad s := -\rho(\gamma^{k+1} - \gamma^k) \qquad for \;\; \varepsilon_s > 0
\end{aligned}$$