Exercises 7: Alternating Direction Method of Multipliers

Jared Fisher

Last Updated October 26, 2016

1 Derivation

Lasso regression invokes the following optimization problem.

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

This can be rewritten into

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}||X\beta - y||_2^2 + \lambda||\gamma||_1 \\ \text{subject to} \quad & \beta - \gamma = 0 \end{aligned}$$

in order to align with ADMM form. Thus, the augmented Lagrangian is

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

This means our Lasso ADMM has the following iterations:

$$\begin{aligned} \beta^{k+1} &:= arg\min_\beta L_\rho(\beta, \gamma^k, v^k) \\ \gamma^{k+1} &:= arg\min_\gamma L_\rho(\beta^{k+1}, \gamma, v^k) \\ v^{k+1} &:= v^k + \rho(\beta^{k+1} - \gamma^{k+1}). \end{aligned}$$

Now to evaluate the argmin's.

$$\begin{aligned} arg\min_\beta L_\rho(\beta, \gamma^k, v^k) &= arg\min_\beta \frac{1}{2}||X\beta - y||_2^2 + v^T\beta + \frac{\rho}{2}||\beta - \gamma||_2^2 \\ \Rightarrow 0 &\overset{set}{=} \frac{\partial}{\partial\beta}\left[\frac{1}{2}||X\beta - y||_2^2 + v^T\beta + \frac{\rho}{2}||\beta - \gamma||_2^2\right] \\ 0 &= X^T(X\beta - y) + v + \rho(\beta - \gamma) \\ 0 &= X^TX\beta - X^Ty + v + \rho\beta - \rho\gamma \\ X^TX\beta + \rho\beta &= X^Ty - v + \rho\gamma \\ (X^TX + \rho I)\beta &= X^Ty - v + \rho\gamma \\ \beta &= (X^TX + \rho I)^{-1}(X^Ty - v + \rho\gamma) \\ arg\min_\beta L_\rho(\beta, \gamma^k, v^k) &= (X^TX + \rho I)^{-1}(X^Ty - v + \rho\gamma) \end{aligned}$$

AND

$$
\begin{aligned}
arg \min_{\gamma} L_\rho(\beta^{k+1}, \gamma, v^k) &= arg \min_{\gamma} \lambda ||\gamma||_1 - v^T\gamma + \frac{\rho}{2}||\beta - \gamma||_2^2 \\
&= arg \min_{\gamma} \lambda ||\gamma||_1 - v^T\gamma + \frac{\rho}{2}\gamma^T\gamma + -\rho\beta^T\gamma \\
&= arg \min_{\gamma} \lambda ||\gamma||_1 - (v^T + \rho\beta^T)\gamma + \frac{\rho}{2}\gamma^T\gamma \\
&= arg \min_{\gamma} \frac{\lambda}{\rho} ||\gamma||_1 - \left(\frac{1}{\rho}v^T + \beta^T\right)\gamma + \frac{1}{2}\gamma^T\gamma \\
&= arg \min_{\gamma} \frac{\lambda}{\rho} ||\gamma||_1 + \frac{1}{2}\left|\left|\gamma - \left(\frac{1}{\rho}v + \beta\right)\right|\right|_2^2 \quad \text{a la proof from ex06} \\
&= S_{\lambda/\rho}\left(\frac{1}{\rho}v + \beta\right)
\end{aligned}
$$

We can also use the scaled augmented Lagrangian to get the same solutions as in the text (`http://stanford.edu/~boyd/papers/pdf/admm_distr_stats.pdf`). Let $u = (1/\rho)v$. Then

$$
\begin{aligned}
\beta^{k+1} &:= (X^TX + \rho I)^{-1}(X^Ty - \rho u + \rho\gamma) \\
&= (X^TX + \rho I)^{-1}(X^Ty + \rho(\gamma - u)) \\
\gamma^{k+1} &:= S_{\lambda/\rho}(u + \beta) \\
v^{k+1} &:= v^k + \beta^{k+1} - \gamma^{k+1}.
\end{aligned}
$$

Lastly, note the stopping rules. For $\epsilon_r, \epsilon_s > 0$,

$$
\begin{aligned}
||r||_2 < \epsilon_r, &\qquad r := \beta - \gamma \\
||s||_2 < \epsilon_s, &\qquad s := -\rho(z^{k+1} - z^k)
\end{aligned}
$$

2 Implementation

The assignment this week is simple: implement ADMM for fitting the lasso regression model, and compare it to your proximal gradient implementation from last week. The application of ADMM to the lasso model is described in Section 6.4 of the Boyd et. al. paper. In the exercises to follow, we'll use ADMM again for several other problems, including spatial smoothing.