Alternating Direction Method of Multipliers

Last week we saw that a problem of interest called the Lasso is written as

$$\min_{w \in \mathbb{R}^D} \|Xw - y\|_2^2 + \lambda \|w\|_1 \qquad (1)$$

which has the more general form

$$\min_{w \in \mathbb{R}^D} f(w) + g(w), \qquad (2)$$

where $f : \mathbb{R}^D \to \mathbb{R}$ and $g : \mathbb{R}^D \to \mathbb{R}$ are convex functions.

It can be difficult and slow to solve problems like (2) using gradient descent. An alternative approach is to use the Alternating Direction Method of Multipliers (ADMM). ADMM applies to optimization problems of the form

$$\begin{aligned} \min_{\substack{x \in \mathbb{R}^p \\ y \in \mathbb{R}^q}} \quad & f(x) + g(y) \qquad (3) \\ \text{ST} \quad & Ax + By = c \end{aligned}$$

"subject to"

where $A \in \mathbb{R}^{r \times p}$, $B \in \mathbb{R}^{r \times q}$, $c \in \mathbb{R}^r$.

The ADMM algorithm proceeds by minimizing the augmented Lagrangian

$$\mathcal{L}_\rho(x,y,\lambda) = \underbrace{f(x)+g(y)+\lambda^T(Ax+By-c)}_{\text{Lagrangian}} + \underbrace{\frac{\rho}{2}\|Ax+By-c\|_2^2}_{\text{augmentation}}$$

We wish to minimize $\mathcal{L}(x,y,\lambda)$ over the three variables x, y, and λ. To do this, ADMM takes an alternating approach (hence the name)

$$\boxed{\begin{array}{l}
\underline{\text{Initialize:}}\ y_0 \in \mathbb{R}^q,\ \lambda_0 \in \mathbb{R}^r \\[1em]
\underline{\text{Iterate:}} \\[1em]
\quad \bullet\ x_{k+1} = \underset{x\in\mathbb{R}^p}{\arg\min}\ \mathcal{L}_\rho(x, y_k, \lambda_k) \\[1em]
\quad \bullet\ y_{k+1} = \underset{y\in\mathbb{R}^q}{\arg\min}\ \mathcal{L}_\rho(x_{k+1}, y, \lambda_k) \\[1em]
\quad \bullet\ \lambda_{k+1} = \lambda^k + \rho\left(Ax_{k+1} + By_{k+1} - c\right)
\end{array}}$$

To make solving for x_{k+1} and y_{k+1} easier, we introduce the proximity or proximal operator

$$\boxed{\mathrm{prox}_f(v) = \underset{x}{\mathrm{argmin}} \; f(x) + \frac{1}{2}\|x - v\|_2^2}$$

To see where this comes in, note that

$$\underset{x}{\mathrm{argmin}} \; f(x) + g(y_k) + \lambda_k^T(Ax + By_k - c) + \frac{\rho}{2}\|Ax + By_k - c\|_2^2$$

$$= \underset{x}{\mathrm{argmin}} \; f(x) + \lambda_k^T Ax + \frac{\rho}{2}\|Ax + By_k - c\|_2^2$$

$$= \underset{x}{\mathrm{argmin}} \; f(x) + \frac{\rho}{2}\|Ax + By_k - c + \frac{1}{\rho}\lambda_k\|_2^2$$

where the third line follows by completing the square.

Let

$$u_k = \frac{1}{\rho}\lambda_k.$$

In the case where $A = I_p$, $B = -I_q$, and $c = 0$, which is common, the iterations become

$$\boxed{\begin{aligned} &\bullet\; x_{k+1} = \mathrm{prox}_{f/\rho}(y_k - u_k) \\ &\bullet\; y_{k+1} = \mathrm{prox}_{g/\rho}(x_{k+1} + u_k) \\ &\bullet\; \lambda_{k+1} = \lambda_k + x_{k+1} - y_{k+1} \end{aligned}}$$

check these for yourself using the definition of the prox operator

Solving the Lasso

The Lasso (1) almost fits the formulation (3) if we let $x = y$. To obtain the solution to (1) using ADMM, we will apply a common technique called variable splitting. To do this, we rewrite the Lasso

$$\min_{w \in \mathbb{R}^D} \frac{1}{2}\|Xw - y\|_2^2 + \lambda \|w\|_1$$

as

$$\min_{\substack{w \in \mathbb{R}^D \\ z \in \mathbb{R}^D}} \frac{1}{2}\|Xw - y\|_2^2 + \lambda \|z\|_1 \qquad (4)$$

$$\text{ST} \qquad w - z = 0$$

which fits the form of (3) but has the same solution as (1). We now determine the values of the proximal operators for f and g.

$$\text{prox}_{f/\rho}(v) = \arg\min_w \frac{1}{2}\|Xw - y\|_2^2 + \frac{\rho}{2}\|w - v\|_2^2$$

$$= \arg\min_w \frac{1}{2}\left(w^T X^T X w + y^T y - 2y^T X w\right) +$$

$$\frac{\rho}{2}\left(w^T w + v^T v - 2w^T v\right)$$

$$= \underset{w}{\operatorname{argmin}} \quad w^T X^T X w + \rho w^T w - 2 y^T X w - 2\rho w^T v$$

$$= \underset{w}{\operatorname{argmin}} \quad w^T (X^T X + \rho I) w - 2 (y^T X + \rho v^T) w$$

$$= \boxed{(X^T X + \rho I)^{-1} (X^T y + \rho v)}$$

$$\operatorname{prox}_{g/\rho}(v) = \underset{z}{\operatorname{argmin}} \quad \lambda \|z\|_1 + \frac{\rho}{2} \|z - v\|_2^2$$

To find the solution, consider first $z \in \mathbb{R}$. In this case

$$\operatorname{prox}_{g/\rho}(v) = \underset{z \in \mathbb{R}}{\operatorname{argmin}} \quad \lambda |z| + \frac{\rho}{2} (z - v)^2$$

Since $|\cdot|$ is not differentiable, we again take the subgradient and set it to zero to minimize. This becomes

$$\partial \left(\lambda |z| + \frac{\rho}{2} (z - v)^2 \right) = \partial \lambda |z| + \rho (z - v) = 0$$

$$\Longleftrightarrow (v - z) = \frac{1}{\rho} \partial \lambda |z|$$

which breaks down into three cases, since

$$\partial \lambda |z| = \begin{cases} \lambda & z > 0 \\ -\lambda & z < 0 \\ [-\lambda, \lambda] & z = 0 \end{cases}$$

is the subdifferential of $\lambda|z|$. If $z>0$, then $\partial\lambda|z|=\lambda$ and we get $z=v-\frac{\lambda}{\rho}$. If $z<0$, $\partial\lambda|z|=-\lambda$ and we get $z=v+\frac{\lambda}{\rho}$. Finally, if $z=0$, then $\partial\lambda|z|=0$, which holds for the range $v\in[-\frac{\lambda}{\rho},\frac{\lambda}{\rho}]$. Summarizing, we see that

$$z=\begin{cases} v-\frac{\lambda}{\rho} & \text{if } v>\frac{\lambda}{\rho} \\ v+\frac{\lambda}{\rho} & \text{if } v<-\frac{\lambda}{\rho} \\ 0 & \text{if } v\in[-\frac{\lambda}{\rho},\frac{\lambda}{\rho}] \end{cases}$$

which is known as the soft thresholding operator $S_{\lambda/\rho}(v)$.

To generalize to multiple dimensions, note that

$$\lambda\|z\|_1+\frac{\rho}{2}\|z-v\|_2^2=\sum_{i=1}^{D}\left(\lambda|z_i|+\frac{\rho}{2}(z_i-v_i)^2\right)$$

so we can solve the proximal operator for each element independently. This yields

$$\boxed{\operatorname{prox}_{g/\rho}(v)=\left(S_{\lambda/\rho}(v_i)\right)_{i=1}^{D}\in\mathbb{R}^D}$$

which for convenience is sometimes written as

$$\boxed{\operatorname{prox}_{g/\rho}(v)=S_{\lambda/\rho}(v).}$$

ADMM for Lasso

Combining the updates above gives the following ADMM iteration for solving the Lasso.

- $w_{k+1} = \mathrm{prox}_{f/\rho}(z_k - u_k)$

$$= (X^T X + \rho I)^{-1} \left(X^T y + \rho (z_k - u_k) \right)$$

- $z_{k+1} = \mathrm{prox}_{g/\rho}(w_{k+1} + u_k)$

$$= S_{\lambda/\rho}\left(w_{k+1} + u_k\right)$$

- $u_{k+1} = u_k + w_{k+1} - z_{k+1}$