GroupLasso

habakan

May 16, 2019

Soft Threshold の導出

ISTA

Group Lasso

Soft Threshold の導出

以下の最適化問題を考える

$$\underset{\boldsymbol{w} \in \mathbb{R}^d}{\mathrm{minimize}} \left(\hat{L}(\boldsymbol{w}) + \lambda \|\boldsymbol{w}\|_1 \right) \tag{1}$$

prox 作用素の定義

$$prox_g(y) = \operatorname*{argmin}_{w \in \Re^d} \left(\frac{1}{2} \|y - w\|_2^2 + g(w) \right) \tag{2}$$

関数 $prox_g(y)$ は強凸関数である.

l_1 ノルムに関する prox 作用素

$$prox_\lambda^{l_1}(y) = \operatorname*{argmin}_{w \in \Re^d} \left(\frac{1}{2} \|y - w\|_2^2 + \|w\|_1 \right) \tag{3}$$

$l_2$prox 作用素を解析的に解を求める.

prox 作用素は強凸関数であるため, 微分をすることにより解析的に求めることができる.

$$\underset{w \in \mathfrak{R}^d}{\operatorname{argmin}} \left(\frac{1}{2} \|y - w\|_2^2 + \|w\|_2 \right) = \frac{1}{1+\lambda} y \tag{4}$$

$l_1$prox 作用素を解析的に解を求める.

$$\frac{1}{2}\|y-w\|_2^2+\|w\|_1=\sum_{j=1}^{d}\left(\frac{1}{2}\|y_j-w_j\|_2^2+\|w_j\|_1\right) \tag{5}$$

和に分解できたため, 各 j について最小化をすることを考える.
最小化を達成する点では w の劣微分は 0 を含むことから

$$y_j-w_j\in\partial\|w_j\|, j=1,2,3,\cdots d \tag{6}$$

絶対値関数 $\|w\|$ の劣微分

$$\partial\|w\| = \begin{cases} -1, & w < 0 \\ [-1, 1], & w = 0 \\ 1, & w > 0 \end{cases} \tag{7}$$

$w = 0$ の時, 劣微分は 1 点には絞ることができない.

$$y_j = [-\lambda, \lambda] \tag{8}$$

w についてそれぞれの場合分けを考える.

$$\left[prox_{\lambda}^{l_1}(y)\right]_j = \begin{cases} y_j + \lambda, & y_j < -\lambda \\ 0, & -\lambda \leq y_j \leq \lambda \\ y_j - \lambda, & y_j > \lambda \end{cases} \tag{9}$$

関数 $prox_{\lambda}^{l_1}$ は Soft Treshold Function と呼ばれる.

ISTA

以下のように w の更新を行う.

$$w^{t+1} = \underset{w}{\operatorname{argmin}} \left(\left\langle \nabla \hat{L}(w^t), w - w^t \right\rangle + \lambda \|w\|_1 - \frac{1}{2\eta_t} \|w - w^t\|_2^2 \right) \quad (10)$$

第一項：$\hat{L}(w)$ を線形近似

第三項：近接項 w^t より値が離れると大きくなる.

第一項と第三項をまとめると更新式は以下の prox 作用素になる.

$$w^{t+1} = prox^{l_1}_{\lambda\eta_t} \left(w^t - \eta_t \nabla \hat{L}(w^t) \right) \quad (11)$$

η_t ：近接項を強くするパラメータ

Group Lasso

グループ l_1 ノルムに関する prox 作用素

$$prox_{\lambda}^{\mathfrak{G}} = \underset{w \in \Re^d}{\operatorname{argmin}} \left(\frac{1}{2} \|y - w\|_2^2 + \lambda \sum_{\mathfrak{g} \in \mathfrak{G}} \|w_{\mathfrak{g}}\|_p \right) \tag{12}$$

グループの重複がないことを仮定

近接項を分解

$$\|y - w\|_2^2 = \sum_{\mathfrak{g} \in \mathfrak{G}} \|y_{\mathfrak{g}} - w_{\mathfrak{g}}\|_2^2 \tag{13}$$

$$y - w \in \lambda \partial \|w\|_2 \tag{14}$$

劣微分 $\partial\|w\|_2$ は原点以外では 1 点 $w/\|w\|_2$ からなる集合,
原点以外では集合 $\{g \in \mathcal{R}^d : \|g\|_2 \leq 1\}$ となる.
$\|y\|_2 < \lambda$ のとき, $w = 0$ で最適となり, それ以外の場合, y を原点方向に長さ λ だけ縮小した点で最適

$$prox_{\lambda\|\cdot\|_2}(y) = \begin{cases} (\|y\|_2 - \lambda)\frac{y}{\|y\|_2}, & \|y\|_2 > \lambda \\ 0, & otherwise \end{cases} \tag{15}$$