Problem

(y_i, x_i), $i = 1 \ldots n$. observed. $y_i \in \mathbb{R}$, $x_i \in \mathbb{R}^p$

W_{ij} — adjacency between i, j

$$\underset{\substack{\{\beta_i\}_{i=1}^n \\ \beta_i \in \mathbb{R}^p}}{\text{minimize}} \quad \underbrace{\sum_{i=1}^{n} \frac{1}{2}\left(y_i - x_i^T \beta_i\right)^2}_{g(\beta)} + \underbrace{\lambda \sum_{i<j} W_{ij} \|\beta_i - \beta_j\|}_{h(\beta)} \quad \text{where } \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \in \mathbb{R}^{np}$$

$$\underset{\beta}{\text{minimize}} \quad g(\beta) + h(\beta)$$

derive the proximal gradient algorithm to solve it assuming the proximal operator P_h is available.

Variable Used:

(y_i, x_i), i = 1,...,n observed

$y_i \in R$

$x_i \in R^p$

w_{ij}: adjacency between i and j, this is an observed value

β_i: node i,

$\beta_i \in R^p \;\text{-}> \beta \in R^{np}$

Optimization problem:

$$min_{\beta_i\, i\, to\, n\, (\beta_i \in R^p)} \frac{1}{2}\sum_{i=1}^{n}(y_i - x_i^T\beta_i)^2 + \lambda\sum_{i<j} w_{ij}\left|\beta_i - \beta_j\right|$$

Formulating proximal gradient descent:

$$g(\beta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - x_i^T\beta_i)^2$$

-> $\frac{\partial g(\beta)}{\partial \beta_i} = (x_i^T\beta_i - y_i)x_i$

-> $\nabla g(\beta)$ = $((x_1^T\beta_1 - y_1)x_1, (x_2^T\beta_2 - y_2)x_2, \ldots, (x_n^T\beta_n - y_n)x_n)^T$

$$h(\beta) = \lambda\sum_{i<j} w_{ij}\left|\beta_i - \beta_j\right|$$

Pseudo code:

Starting from an initial value β^0, iterate for k = 1, 2,…,n

t = 1
while True:

$$\nabla g(\beta^{k-1}) = ((x_1^T\beta^{k-1}{}_1 - y_1)x_1, (x_2^T\beta^{k-1}{}_2 - y_2)x_2, \ldots, (x_n^T\beta^{k-1}{}_n - y_n)x_n)^T$$

$$G_{t_k}(\beta^{k-1}) = \frac{\beta^{k-1} - \mathrm{P}_{h,t}(\beta^{k-1} - t_k\nabla g(\beta^{k-1}))}{t}$$

$$\beta^k = \beta^{k-1}G_{t_k}(\beta^{k-1})$$

if $g(\beta^k - t_kG_{t_k}(\beta^{k-1})) \le g(\beta^{k-1}) - t\nabla g(\beta^{k-1})^TG_t(\beta^{k-1}) + \frac{t}{2}\left|G_{t_k}(\beta^{k-1})\right|^2$

break

else:

$t_k = t_k\alpha$

Note:

$\mathrm{P}_{h,t}(\beta^{k-1} - t_k\nabla g(\beta^{k-1}))$

$$= argmin_u \frac{1}{2}\left|u - (\beta^{k-1} - t_k \nabla g(\beta^{k-1}))\right|^2 + th(u)$$

$$= argmin_u \frac{1}{2}\left|u - (\beta^{k-1} - t_k \nabla g(\beta^{k-1}))\right|^2 + t\lambda \sum_{i<j}\left|u_i - u_j\right|$$

$$= argmin_u \frac{1}{2}\Sigma^n_{i=1}\left|u_i - w_i\right|^2 + t\lambda \sum_{i<j}\left|u_i - u_j\right|$$

$= f(w, \lambda t)$ -> returns minimized u

Question:

How to solve the Proximal operator?
https://arxiv.org/abs/1507.00280 - Hallac's
https://arxiv.org/abs/1908.02370 - Improved algorithm
https://github.com/alexfengg/novelGFL/tree/master/python - source code