Network + L_2-norm for DNA-methylation data.

$$y_i = \beta_0 + \sum \beta_j \cdot x_{ij} + \varepsilon_i$$

$$\underbrace{\frac{1}{n}\| y - X\beta \|_2^2}_{\text{or } l(y, X\beta)} + \lambda_1 \cdot \|\beta\|_1 + \frac{\lambda_2}{2} \cdot \beta^T L \beta$$

$$= l(y, X\beta) + \lambda_1 \cdot \|\beta\|_1 + \frac{\lambda_2}{2} \cdot \sum_{u \sim v} w_{uv} \cdot \left\{ \frac{\beta_u}{\sqrt{d_u}} - \frac{\beta_v}{\sqrt{d_v}} \right\}$$

$$\boxed{}$$

$$y_i = \beta_0 + \sum_{k=1}^{d} \underset{\sim}{x}_i^{(k)T} \cdot \beta^{(k)} + \varepsilon_i, \quad d \text{ groups. (gene)}$$

$$\underset{\sim}{X}_i = [x_i^{(1)T}, \cdots, x_i^{(d)T}] = \begin{bmatrix} x_i^{(1)} \\ \vdots \\ x_i^{(d)} \end{bmatrix}$$

$$l(y, X\beta) + \lambda_1 \cdot \|\beta\|_1 + \frac{\lambda_2}{2} \underset{p\times p}{\beta^T} \overset{d\times d}{\underset{p\times p}{L}} \beta$$

$$\Longrightarrow l(y, X\beta) + \lambda_1 \cdot \|\beta\|_1 + \frac{\lambda_2}{2} \cdot \underset{1\times d}{\hat{\beta}^T} \underset{d\times d}{L} \underset{d\times 1}{\hat{\beta}}$$

where $\hat{\beta} = (\|\beta^{(1)}\|_2, \cdots, \|\beta^{(d)}\|_2)^T \in \mathbb{R}^{d\times 1}$

$$= l(y, X\beta) + \lambda_1 \cdot \|\beta\|_1 + \frac{\lambda_2}{2} \cdot \sum_{u \sim v} \left(\frac{\hat{\beta}_u}{\sqrt{d_u}} - \frac{\hat{\beta}_v}{\sqrt{d_v}} \right)$$

$$= l(y, X\beta) + \lambda_1 \cdot \|\beta\|_1 + \frac{\lambda_2}{2} \cdot \sum_{u \sim v} \left(\frac{\|\beta_u\|_2}{\sqrt{d_u}} - \frac{\|\beta_v\|_2}{\sqrt{d_v}} \right)$$

When $l(y, X\beta) = \frac{1}{n}\| y - X\beta \|_2^2$,

$$L(\beta; X, y) = \frac{1}{n}\| y - X\beta \|_2^2 + \lambda_1 \cdot \|\beta\|_1 + \frac{\lambda_2}{2} \cdot \sum_{u \sim v} \left\{ \frac{\|\beta_u\|_2}{\sqrt{d_u}} - \frac{\|\beta_v\|_2}{\sqrt{d_v}} \right\}$$

$$\frac{\partial L}{\partial \beta_u} = -\frac{1}{n} X_u^T \underbrace{(y - X\beta)}_{= y - X_u \beta_u - X_{-u} \beta_{-u}} + \lambda_1 \cdot \text{sgn}(\underset{\sim}{\beta}_u) + \frac{\lambda_2}{2} \cdot \frac{d_u}{\sqrt{d_u}} \left\{ \frac{\beta_u}{\|\beta_u\|_2} \right\} = 0$$

for β : $\|\beta\|_2 \neq 0$

$$= -\frac{1}{n} X_u^T y + \frac{1}{n} X_u^T X_{-u} \cdot \beta_{-u} + \frac{1}{n} X_u^T X_u \beta_u + \lambda_1 \cdot \text{sgn}(\underset{\sim}{\beta}_u) + \frac{\lambda_2}{2} \cdot \sqrt{d_u} \frac{\beta_u}{\|\beta_u\|_2} = 0$$

$$\Big(\frac{1}{n} \underset{d_u\times n}{X_u^T} \underset{n\times d_u}{X_u} + \frac{\lambda_2}{2} \cdot \frac{\sqrt{d_u}}{\|\beta_u\|_2} I_{d_u} \Big) \cdot \underset{d_u \times 1}{\hat{\beta}_u} = \left\{ \frac{1}{n} X_u^T (y - X_{-u} \beta_{-u}) - \lambda_1 \cdot \text{sgn}(\underset{\sim}{\beta}_u) \right\}$$

If $n > d_u$

$$\underset{\sim}{\beta}_u = \left\{ \frac{1}{n} X_u^T X_u + \frac{\lambda_2}{2} \cdot \frac{\sqrt{d_u}}{\|\beta_u\|_2} I_{d_u} \right\}^{-1} \cdot \Big\{ \underbrace{\frac{1}{n} X_u^T (\underset{\wedge}{y} - X_{-u} \beta_{-u})}_{\triangleq Z_u} - \lambda_1 \cdot \text{sgn}(\underset{\sim}{\beta}_u) \Big\}$$

$$\|\underset{\sim}{\beta}_u\|_2 = \left\{ (Z_u - \lambda_1 \text{sgn}(\beta_u))^T \cdot \left(\frac{1}{n} X_u^T X_u + \frac{\lambda_2}{2} \cdot \frac{\sqrt{d_u}}{\|\beta_u\|_2} I_{d_u} \right)^{-1} \cdot \left(\frac{1}{n} X_u^T X_u + \frac{\lambda_2}{2} \frac{\sqrt{d_u}}{\|\beta_u\|_2} I_{d_u} \right)^{-1} (Z_u - \lambda_1 \text{sgn}(\beta_u)) \right\}$$

Quadratic Majorization

$$L(\beta|D) \le L(\hat{\beta}|D) + (\beta-\hat{\beta})^T \nabla L(\hat{\beta}|D) + \frac{1}{2}(\beta-\hat{\beta})^T H (\beta-\hat{\beta}),$$

where $\nabla L(\hat{\beta}|D) = -X^T(y - X\hat{\beta})/n$.

$$l(y, X\beta) + \lambda_1 \cdot \sum_{k=1}^{d} \|\beta^{(k)}\|_2 + \lambda_2 \cdot \sum_{u\sim v} \left\{ \frac{\check{\beta}_u}{\sqrt{d_u}} - \frac{\check{\beta}_v}{\sqrt{d_v}} \right\}^2, \quad \check{\beta} = \begin{bmatrix} \|\beta^{(1)}\|_2 \\ \vdots \\ \|\beta^{(d)}\|_2 \end{bmatrix}$$

$$Q(\beta|D) = l(y, X\beta) + \lambda_1 \cdot \sum_{k=1}^{d} \|\beta^{(k)}\|_2 + \lambda_2 \cdot \sum_{u\sim v} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\}^2 \qquad \check{\beta}_u = \|\beta^{(u)}\|_2$$

when $l(y, X\beta) = \frac{1}{n}\|y - X\beta\|_2^2$

$$\le L(\bar{\beta}|D) + (\beta-\bar{\beta})^T \nabla L(\bar{\beta}|D) + \frac{1}{2}(\beta-\bar{\beta})^T H (\beta-\bar{\beta})$$

β: new

$\bar{\beta}$: old

$$+ \lambda_1 \sum_{k=1}^{d} \|\beta^{(k)}\|_2 + \lambda_2 \cdot \sum_{u\sim v} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\}^2$$

$$\mathcal{L}^*(\beta - \bar{\beta})$$

$$\frac{\partial}{\partial \beta^{(u)}} \left\{ \sum_{u\sim v} \left(\frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right)^2 \right\} = \sum_{v\sim u} 2 \cdot \left(\frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right) \cdot \frac{1}{\sqrt{d_u}} \cdot \frac{\beta^{(u)}}{\|\beta^{(u)}\|_2} \quad \underline{\text{if } \|\beta^{(u)}\|_2 \ne 0}$$

$$= 2\cdot\beta^{(u)} - \sum_{v\sim u} 2 \cdot \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}} \cdot \frac{\beta^{(u)}}{\|\beta^{(u)}\|_2}$$

$$= 2\cdot\beta^{(u)} \cdot \left\{ 1 - \frac{1}{\sqrt{d_u} \cdot \|\beta^{(u)}\|_2} \cdot \sum_{v\sim u} \frac{\|\beta^{(v)}\|}{\sqrt{d_v}} \right\}$$

$$(\beta - \bar{\beta}) = (0, \cdots, 0, \beta^{(u)} - \bar{\beta}^{(u)}, 0, \cdots, 0)^T$$

$$\mathcal{L}^*(\beta^{(u)} - \bar{\beta}^{(u)}) = L(\bar{\beta}^{(u)}|D) + (\beta^{(u)} - \bar{\beta}^{(u)})^T \nabla L(\bar{\beta}|D)^{(u)} + \frac{1}{2}(\beta^{(u)} - \bar{\beta}^{(u)})^T \cdot H^{(u)} \cdot (\beta^{(u)} - \bar{\beta}^{(u)})$$

$|N_u| = d_u$

$N_u = \{v : v \sim u\}$

* $H^{(u)} = \frac{1}{n} \underset{d_u \times n}{X^{(u)T}} \underset{n \times d_u}{X^{(u)}}$ is the submatrix of $H = \frac{1}{n}X^TX$ for the u-th group.

* $\nabla L(\bar{\beta}|D)^{(u)} = -\frac{1}{n} X^{(u)T}(y - X\bar{\beta})$.

We minimize
$$L(\bar{\beta}^{(u)}|D) + (\beta^{(u)} - \bar{\beta}^{(u)})^T \cdot \nabla L(\bar{\beta}|D)^{(u)} + \frac{1}{2}(\beta^{(u)} - \bar{\beta}^{(u)})^T H^{(u)} (\beta^{(u)} - \bar{\beta}^{(u)}) + \lambda_1 \cdot \|\beta^{(u)}\|_2 + \frac{\lambda_2}{2} \cdot \sum_{v\sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\}^2$$

$$\le L(\bar{\beta}^{(u)}|D) + (\beta^{(u)} - \bar{\beta}^{(u)})^T \cdot \nabla L(\bar{\beta}|D)^{(u)} + \frac{\gamma_u}{2}(\beta^{(u)} - \bar{\beta}^{(u)})^T (\beta^{(u)} - \bar{\beta}^{(u)}) + \lambda_1 \cdot \|\beta^{(u)}\|_2 + \frac{\lambda_2}{2} \cdot \sum_{v\sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\}^2$$

$$= R(\beta^{(u)})$$

When $\|\beta^{(u)}\|_2 \ne 0$,

$$\frac{\partial R(\beta^{(u)})}{\partial \beta^{(u)}} = \nabla L(\bar{\beta}|D)^{(u)} + \gamma_u(\beta^{(u)} - \bar{\beta}^{(u)}) + \lambda_1 \cdot \frac{\beta^{(u)}}{\|\beta^{(u)}\|_2} + \lambda_2 \cdot \beta^{(u)} \cdot \left\{ 1 - \frac{1}{\|\beta^{(u)}\|_2} \cdot \sum_{v\sim u} \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u \cdot d_v}} \right\} = 0$$

$$= \nabla L(\bar{\beta}|D)^{(u)} - \gamma_u \cdot \bar{\beta}^{(u)} + \beta^{(u)} \cdot \underbrace{\left\{ \gamma_u + \lambda_1 \cdot \frac{1}{\|\beta^{(u)}\|_2} + \lambda_2 - \frac{\lambda_2}{\|\beta^{(u)}\|_2} \cdot \sum_{v\sim u} \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}} \right\}}_{= \gamma_u + \lambda_2 + \frac{1}{\|\beta^{(u)}\|_2} \cdot (\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u})} = 0$$

where $\|\beta^{(v)}\|_{v\sim u} = \sum_{v\sim u} \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}}$

$$\beta^{(u)} = \left\{ \gamma_u + \lambda_2 + \frac{\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u}}{\|\beta^{(u)}\|_2} \right\}^{-1} \cdot \left\{ \gamma_u \cdot \bar{\beta}^{(u)} - \nabla L(\bar{\beta}|D)^{(u)} \right\}$$

$$\|\beta^{(u)}\|_2 = \left\{ \gamma_u + \lambda_2 + \frac{\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u}}{\|\beta^{(u)}\|_2} \right\}^{-1} \cdot \| \gamma_u \cdot \bar{\beta}^{(u)} - \nabla L(\bar{\beta}|D)^{(u)} \|_2$$

$$(\gamma_u + \lambda_2) \cdot \|\beta^{(u)}\|_2 + (\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u}) = \| \gamma_u \cdot \bar{\beta}^{(u)} - \nabla L(\bar{\beta}|D)^{(u)} \|_2$$

$$\therefore \|\beta^{(u)}\|_2 = \frac{1}{(\gamma_u + \lambda_2)} \cdot \left\{ \| \gamma_u \cdot \bar{\beta}^{(u)} - \nabla L(\bar{\beta}|D)^{(u)} \|_2 - (\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u}) \right\}$$

$$\therefore \therefore \beta^{(u)} = \left\{ \gamma_u + \lambda_2 + \frac{(\gamma_u + \lambda_2) \cdot (\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u})}{\{ \| \gamma_u \cdot \bar{\beta}^{(u)} - \nabla L(\bar{\beta}|D)^{(u)} \|_2 - (\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u}) \}} \right\}^{-1}$$

$$\times \left\{ \gamma_u \cdot \bar{\beta}^{(u)} - \nabla L(\bar{\beta}|D)^{(u)} \right\}$$

$$= \frac{1}{\gamma_u + \lambda_2} \cdot \left\{ 1 + \frac{(\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u})}{\{ \|W^{(u)}\|_2 - (\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u}) \}} \right\}^{-1} \times W^{(u)}$$

$$= \frac{1}{\gamma_u + \lambda_2} \times \left\{ \|W^{(u)}\|_2 - (\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u}) \right\} \frac{W^{(u)}}{\|W^{(u)}\|_2}$$

$$\nabla L(\bar{\beta}|D)^{(u)} = -\frac{1}{n} X^{(u)T}(y - X\bar{\beta})$$

where $W^{(u)} = \gamma_u \cdot \bar{\beta}^{(u)} - \nabla L(\bar{\beta}|D)^{(u)}$

$$\therefore \beta^{(u)}_{new} = \frac{1}{\gamma_u + \lambda_2} \times \left\{ 1 - \frac{(\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u})}{\|W^{(u)}\|_2} \right\} \cdot W^{(u)}, \qquad W^{(u)} = \gamma_u \cdot \beta^{(u)}_{old} + \frac{1}{n} X^{(u)T}(y - X\beta_{old})$$

Quadratic Majorization

$$L(\beta|D) \le L(\hat\beta|D) + (\beta-\hat\beta)^T \nabla L(\hat\beta|D) + \frac{1}{2}(\beta-\hat\beta)^T H (\beta-\hat\beta),$$

where $\nabla L(\hat\beta|D) = -X^T(y - X\hat\beta)/n$.

$$l(y, X\beta) + \lambda_1 \cdot \sum_{k=1}^{d} \|\beta^{(k)}\|_2 + \lambda_2 \cdot \sum_{u\sim v} \left\{ \frac{\hat\beta_u}{\sqrt{d_u}} - \frac{\hat\beta_v}{\sqrt{d_v}} \right\}^2, \quad \hat\beta = \begin{bmatrix} \|\beta^{(1)}\|_2 \\ \vdots \\ \|\beta^{(d)}\|_2 \end{bmatrix}$$

$$Q(\beta|D) = l(y, X\beta) + \lambda_1 \cdot \sum_{k=1}^{d} \|\beta^{(k)}\|_2 + \lambda_2 \cdot \sum_{u\sim v} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\}^2 \qquad \hat\beta_u = \|\beta^{(u)}\|_2$$

when $l(y, X\beta) = \frac{1}{n}\|y - X\beta\|_2^2$

$$\le \underline{L(\bar\beta|D) + (\beta-\bar\beta)^T \nabla L(\bar\beta|D) + \frac{1}{2}(\beta-\bar\beta)^T H (\beta-\bar\beta)}$$

β: new
$\bar\beta$: old

$$+ \lambda_1 \cdot \sum_{k=1}^{d} \|\beta^{(k)}\|_2 + \lambda_2 \cdot \sum_{u\sim v} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\}^2 \Big\downarrow$$

$$(\beta-\bar\beta) = (0, \cdots, 0, \beta^{(u)} - \bar\beta^{(u)}, 0, \cdots, 0)^T \qquad \mathcal{L}^*(\beta - \bar\beta)$$

$$\frac{\partial}{\partial \beta^{(u)}} \left\{ \sum_{u\sim v} \left(\frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right)^2 \right\} = \sum_{v\sim u} 2 \cdot \left(\frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right) \cdot \frac{1}{\sqrt{d_u}} \cdot \frac{\beta^{(u)}}{\|\beta^{(u)}\|_2}$$

$$= 2 \cdot \beta^{(u)} - \sum_{v\sim u} 2 \cdot \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}} \cdot \frac{\beta^{(u)}}{\|\beta^{(u)}\|_2} \qquad \underline{\text{if } \|\beta^{(u)}\|_2 \ne 0}$$

$$= 2 \cdot \beta^{(u)} \left\{ 1 - \frac{1}{\sqrt{d_u} \cdot \|\beta^{(u)}\|_2} \cdot \sum_{v\sim u} \frac{\|\beta^{(v)}\|}{\sqrt{d_v}} \right\}$$

$$\mathcal{L}^*(\beta^{(u)} - \bar\beta^{(u)}) = L(\bar\beta^{(u)}|D) + (\beta^{(u)} - \bar\beta^{(u)})^T \nabla L(\bar\beta|D)^{(u)} + \frac{1}{2}(\beta^{(u)} - \bar\beta^{(u)})^T \cdot H^{(u)} \cdot (\beta^{(u)} - \bar\beta^{(u)})$$

$|N_u| = d_u$

$N_u = \{v : v \sim u\}$

* $H^{(u)} = \frac{1}{n} \underset{d_u \times n}{X^{(u)T}} \underset{n \times d_u}{X^{(u)}}$ is the submatrix of $H = \frac{1}{n}X^TX$ for the u-th group.

* $\nabla L(\bar\beta|D)^{(u)} = -\frac{1}{n} X^{(u)T}(y - X\bar\beta)$.

We minimize $L(\bar\beta^{(u)}|D) + (\beta^{(u)} - \bar\beta^{(u)})^T \cdot \nabla L(\bar\beta|D)^{(u)} + \frac{1}{2}(\beta^{(u)} - \bar\beta^{(u)})^T H^{(u)} (\beta^{(u)} - \bar\beta^{(u)}) + \lambda_1 \cdot \|\beta^{(u)}\|_2 + \frac{\lambda_2}{2} \cdot \sum_{v\sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\}^2$

$$\le L(\bar\beta^{(u)}|D) + (\beta^{(u)} - \bar\beta^{(u)})^T \cdot \nabla L(\bar\beta|D)^{(u)} + \frac{\gamma_u}{2}(\beta^{(u)} - \bar\beta^{(u)})^T (\beta^{(u)} - \bar\beta^{(u)}) + \lambda_1 \cdot \|\beta^{(u)}\|_2 + \frac{\lambda_2}{2} \cdot \sum_{v\sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\}^2$$

$$= R(\beta^{(u)})$$

The subgradient equation becomes

$$\frac{\partial R(\beta^{(u)})}{\partial \beta^{(u)}} = \nabla L(\bar\beta|D)^{(u)} + \gamma_u(\beta^{(u)} - \bar\beta^{(u)}) + \lambda_1 \cdot S^{(u)} + \frac{\lambda_2}{2} \cdot \sum_{v\sim u} 2 \cdot \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\} \cdot \frac{S^{(u)}}{\sqrt{d_u}} = 0 \quad \text{for } u = 1, 2, \cdots, d$$

$$S^{(u)} = \begin{cases} \{\beta^{(u)} : \|\beta^{(u)}\|_2 < 1\}, & \text{if } \|\beta^{(u)}\|_2 = 0 \\ \dfrac{\beta^{(u)}}{\|\beta^{(u)}\|_2}, & \text{if } \|\beta^{(u)}\|_2 \ne 0. \end{cases}$$

When $\|\beta^{(u)}\|_2 = 0$, the subgradient equation becomes

$$\nabla L(\bar\beta|D)^{(u)} - \gamma_u \cdot \bar\beta^{(u)} + \lambda_1 \cdot S^{(u)} + \lambda_2 \left(\sum_{v\sim u} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}} \right) S^{(u)} = 0$$

$$\left\{ \lambda_1 - \lambda_2 \cdot \sum_{v\sim u} \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}} \right\} \cdot S^{(u)} = \left\{ -\nabla L(\bar\beta|D)^{(u)} + \gamma_u \bar\beta^{(u)} \right\}$$

$$\therefore S^{(u)} = \left\{ \lambda_1 - \lambda_2 \cdot \sum_{v\sim u} \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}} \right\}^{-1} \cdot \left\{ -\nabla L(\bar\beta|D)^{(u)} + \gamma_u \bar\beta^{(u)} \right\}$$

If $\|S^{(u)}\|_2 = \left\{ \lambda_1 - \lambda_2 \cdot \sum_{v\sim u} \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}} \right\}^{-1} \cdot \| -\nabla L(\bar\beta|D)^{(u)} + \gamma_u \bar\beta^{(u)} \|_2 < 1, \quad \underset{\sim}{\beta^{(u)} = 0}$

$$= \underbrace{\| -\nabla L(\bar\beta|D)^{(u)} + \gamma_u \bar\beta^{(u)} \|_2}_{= \|W\|_2} < \underbrace{\lambda_1 - \lambda_2 \cdot \sum_{v\sim u} \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}}}_{= \lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u}}$$

when $\|\beta^{(u)}\|_2 \neq 0$

① $\therefore \beta^{(u)}_{new} = \frac{1}{\gamma_u + \lambda_2} \times \left\{ 1 - \frac{(\lambda_1 - \lambda_2 \cdot \|\beta^{(v)}\|_{v\sim u})}{\|w^{(u)}\|_2} \right\} \cdot w^{(u)}$, $w^{(u)} = \gamma_u \cdot \beta^{(u)}_{old} + \frac{1}{n} X^{(u)T}(y - X\beta_{old})$

when $\|\beta^{(u)}\|_2 = 0$

If $\|S^{(u)}\|_2 = \left\{ \lambda_1 - \lambda_2 \cdot \sum_{v\sim u} \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}} \right\}^{-1} \cdot \| -\nabla L(\bar{\beta}|D)^{(u)} + \gamma_u \bar{\beta}^{(u)} \|_2 < 1$, $\beta^{(u)} = \underset{\smile}{0}$

② $= \underbrace{\| -\nabla L(\bar{\beta}|D)^{(u)} + \gamma_u \bar{\beta}^{(u)} \|_2}_{= \|W\|_2} < \lambda_1 - \lambda_2 \cdot \sum_{v\sim u} \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u d_v}} \Rightarrow \beta^{(u)} = \underset{\smile}{0}$

①+② : $\beta^{(u)}_{new} = \frac{1}{\gamma_u + \lambda_2} \times \left\{ 1 - \frac{\lambda_1 - \lambda_2 \cdot \|\beta^{(u)}\|_{v\sim u}}{\|w^{(u)}\|_2} \right\}_+ \cdot w^{(u)}$,

where $w^{(u)} = \gamma_u \cdot \beta^{(u)}_{old} + \frac{1}{n} X^{(u)T}(y - X\beta_{old})$

$\{\cdot\}_+ = \max\{0, \cdot\}$

Network + Sparse group lasso $\qquad \lambda_2 \cdot \hat{\beta}^T L \hat{\beta}$

$$\underset{\beta}{\text{minimize}} \ \frac{1}{n}\|y - X\beta\|_2^2 + \lambda_1 \cdot \|\beta\|_1 + \lambda_2 \cdot \sum_{u \sim v} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_u}} \right\}^2$$

$$\leq L(\bar{\beta}|D)^{(u)} + (\beta^{(u)} - \bar{\beta}^{(u)})^T \cdot \nabla L(\bar{\beta}|D)^{(u)} + \frac{\gamma_u}{2}\|\beta^{(u)} - \bar{\beta}^{(u)}\|_2^2 + \lambda_1 \cdot \|\beta^{(u)}\|_1 + \frac{\lambda_2}{2} \sum_{u \sim v} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\}^2$$

$$= \underline{\frac{1}{n}\|y - X\bar{\beta}\|_2^2 + (\beta^{(u)} - \bar{\beta}^{(u)})^T \cdot \left(-\frac{1}{n} X^{(u)T}(y - X\bar{\beta})\right) + \frac{\gamma_u}{2}\|\beta^{(u)} - \bar{\beta}^{(u)}\|_2^2 + \lambda_1 \cdot \|\beta^{(u)}\|_1 + \frac{\lambda_2}{2} \cdot \sum_{u \sim v}} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\}^2$$

$$= Q(\beta^{(u)} | \bar{\beta}^{(u)}, D)$$

$$\frac{\partial Q}{\partial \beta^{(u)}} = \nabla L(\bar{\beta}|D)^{(u)} + \gamma_u \cdot (\beta^{(u)} - \bar{\beta}^{(u)}) + \lambda_1 \cdot \underset{\sim}{S}^{(u)} + \lambda_2 \cdot \sum_{v \sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\} \cdot \frac{1}{\sqrt{d_u}} \times t^{(u)} = 0, \qquad S^{(u)} = \begin{bmatrix} S_1^{(u)} \\ \vdots \\ S_{p_u}^{(u)} \end{bmatrix}$$

$$= -\frac{1}{n} X^{(u)T}(y - X\bar{\beta}) + \gamma_u \cdot (\beta^{(u)} - \bar{\beta}^{(u)}) + \lambda_1 \cdot S^{(u)} + \lambda_2 \cdot \sum_{v \sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\} \cdot \frac{1}{\sqrt{d_u}} \times t^{(u)} = 0$$

$$\frac{\partial Q}{\partial \beta_j^{(u)}} = \nabla L(\bar{\beta}|D)_j^{(u)} + \gamma_u \cdot (\beta_j^{(u)} - \bar{\beta}_j^{(u)}) + \lambda_1 \cdot S_j^{(u)} + \lambda_2 \cdot \sum_{v \sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\} \cdot \frac{1}{\sqrt{d_u}} \times t_j^{(u)} = 0$$

Then, a neccessary and sufficient condition for $\beta^{(u)}$ to be zero is that

$$-\nabla L(\bar{\beta}|D)_j^{(u)} - \gamma_u(\beta_j^{(u)} - \bar{\beta}_j^{(u)}) = \lambda_1 \cdot S_j^{(u)} + \lambda_2 \cdot \sum_{v \sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\} \cdot \frac{1}{\sqrt{d_v}} \times t_j^{(u)}$$

have a solution $\|t^{(u)}\|_2 \leq 1$ and $S_j^{(u)} \in [-1, 1]$. We can determine this by minimizing

$$J(S) = \underbrace{\left\{ \lambda_2 \cdot \sum_{v \sim u} \left(\frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right) \cdot \frac{1}{\sqrt{d_v}} \right\}^{-2}}_{\triangleq \delta} \cdot \sum_{j=1}^{p_u} \left\{ -\lambda_1 \cdot S_j^{(u)} - \nabla L(\bar{\beta}|D)_j^{(u)} - \gamma_u(\beta_j^{(u)} - \bar{\beta}_j^{(u)}) \right\}^2 = \sum_{j=1}^{p_u} (t_j^{(u)})^2$$

with respect to $S_j^{(u)} \in [-1, 1]$. By checking if $J(S) \leq 1$, we have

$$\frac{1}{\delta^2} \sum_j \left\{ -\lambda_1 S_j^{(u)} - \nabla L(\bar{\beta}|D)_j^{(u)} - \gamma_u(\beta_j^{(u)} - \bar{\beta}_j^{(u)}) \right\}^2 = \sum_j \left\{ t_j^{(u)} \right\}^2$$

The minimizer is easily seen to be

$$\hat{S}_j^{(u)} = \begin{cases} -\frac{1}{\lambda_1} \times \left\{ \nabla L(\bar{\beta}|D)_j^{(u)} + \gamma_u(\beta_j^{(u)} - \bar{\beta}_j^{(u)}) \right\} & \text{if } \left| \frac{\nabla L(\bar{\beta}|D)_j^{(u)} + \gamma_u(\beta_j^{(u)} - \bar{\beta}_j^{(u)})}{\lambda_1} \right| \leq 1 \\ -\text{sign}\left\{ \frac{1}{\lambda_1} \cdot \left(\nabla L(\bar{\beta}|D)_j^{(u)} + \gamma_u(\beta_j^{(u)} - \bar{\beta}_j^{(u)}) \right) \right\} & \text{if } \left| \quad '' \quad \right| > 1 \end{cases}$$

The subgradient equation of $Q(\beta^{(u)} | \bar{\beta}^{(u)}, D)$ with respect to $\beta_j^{(u)}$ leads that

$$= -\frac{1}{n} \cdot X^{(u)T} \cdot (y - X\bar{\beta}) + \gamma_u(\beta^{(u)} - \bar{\beta}^{(u)}) + \lambda_1 \cdot S_j^{(u)} + \lambda_2 \cdot \sum_{v \sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\} \cdot \frac{1}{\sqrt{d_u}} \cdot t^{(u)} = 0$$

$$= \lambda_1 \cdot S_j^{(u)} + \lambda_2 \cdot \sum_{v \sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\} \cdot \frac{t^{(u)}}{\sqrt{d_u}} = \frac{1}{n} X^{(u)T}(y - X\bar{\beta}) - \gamma_u(\beta^{(u)} - \bar{\beta}^{(u)}).$$

$$t^{(u)} = \begin{cases} \frac{\beta^{(u)}}{\|\beta^{(u)}\|_2}, & \text{if } \beta^{(u)} \neq 0 \\ \in \left\{ \beta^{(u)} : \|\beta^{(u)}\|_2 \leq 1 \right\}, & \text{if } \beta^{(u)} = \underset{\sim}{0} \end{cases}$$

$$S_j^{(u)} = \begin{cases} \text{sign}(\beta_j^{(u)}), & \text{if } \beta_j^{(u)} \neq 0 \\ \in \left\{ S_j^{(u)} : |S_j^{(u)}| \leq 1 \right\}, & \text{if } \beta_j^{(u)} = 0 \end{cases}$$

where $t^{(u)}$ and $\underset{(S_1^{(u)}, \cdots, S_{p_u}^{(u)})^T}{\overset{\parallel}{S^{(u)}}}$ are subgradients of $\|\beta^{(u)}\|_2$ and $\|\beta^{(u)}\|_1$

We see that the subgradient equations are satisfied with $\beta_j^{(u)} = \underset{\sim}{0}$ if

from $\lambda_1 \cdot S_j^{(u)} - \lambda_2 \cdot \|\beta^{(v)}\|_{v \sim u} \cdot t^{(u)} = \frac{1}{n} X^{(u)T}(y - X\bar{\beta}) + \gamma_u \bar{\beta}^{(u)}$

$$\Rightarrow S_j^{(u)} = \frac{1}{\lambda_1} \cdot \left\{ \lambda_2 \cdot \|\beta^{(v)}\|_{v \sim u} \cdot \frac{\beta_j^{(u)}}{\|\beta^{(u)}\|_2} + \left(\frac{1}{n} X^{(u)T}(y - X\bar{\beta}) + \gamma_u \cdot \bar{\beta}^{(u)} \right) \right\}$$

$$\frac{\partial Q}{\partial \beta^{(u)}} = \nabla L(\bar{\beta}|D)^{(u)} + \gamma_u \cdot \left(\beta^{(u)} - \bar{\beta}^{(u)}\right) + \lambda_1 \cdot \underset{\sim}{S}^{(u)} + \lambda_2 \cdot \sum_{v \sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\} \cdot \frac{1}{\sqrt{d_u}} \times t^{(u)} = 0$$

$$= -\frac{1}{n} X^{(u)T}(y - X\bar{\beta}) + \gamma_u \cdot (\beta^{(u)} - \bar{\beta}^{(u)}) + \lambda_1 \cdot S^{(u)} + \lambda_2 \cdot \sum_{v \sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\} \cdot \frac{1}{\sqrt{d_u}} \times t^{(u)} = 0$$

$$\frac{\partial Q}{\partial \beta_j^{(u)}} = \nabla L(\bar{\beta}|D)_j^{(u)} + \gamma_u \cdot (\beta_j^{(u)} - \bar{\beta}_j^{(u)}) + \lambda_1 \cdot S_j^{(u)} + \lambda_2 \cdot \sum_{v \sim u} \left\{ \frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}} \right\} \cdot \frac{1}{\sqrt{d_u}} \times t_j^{(u)} = 0$$

If $\beta^{(u)} = \underset{\sim}{0}$, $\quad -\frac{1}{n} X^{(u)T}(y - X\bar{\beta}) - \gamma_u \bar{\beta}^{(u)} + \lambda_1 \cdot S^{(u)} - \lambda_2 \cdot \|\beta^{(v)}\|_{v \sim u} \cdot t^{(u)} = 0 \quad \|t^{(u)}\|_2 \le 1.$

$$\|t^{(u)}\|_2 = \left\{\lambda_2 \cdot \|\beta^{(v)}\|_{v \sim u}\right\}^{-1} \cdot \Big\| \underbrace{-\frac{1}{n} X^{(u)T}(y - X\bar{\beta}) - \gamma_u \cdot \bar{\beta}^{(u)}}_{= -W} + \lambda \cdot S^{(u)} \Big\|_2 \le 1$$

$$\Rightarrow \beta^{(u)} = 0.$$

$$\frac{1}{n}\|y-X\beta\|_2^2+\lambda_1\|\beta\|_1+\lambda_2\sum_{k=1}^{d}\|\beta^{(k)}\|_2+\lambda_3\cdot\sum_{u\sim v}\left\{\frac{\|\beta^{(u)}\|_2}{\sqrt{d_u}}-\frac{\|\beta^{(v)}\|_2}{\sqrt{d_v}}\right\}^2$$

$$\leq L(\bar\beta\,|D)+(\beta-\bar\beta)^T\nabla L(\bar\beta\,|D)+\frac{1}{2}(\beta-\bar\beta)^T H(\beta-\bar\beta)+\lambda_1\|\beta\|_1+\lambda_2\cdot\sum_{k=1}^{d}\|\beta^{(k)}\|_2+\lambda_3\cdot\hat\beta^T L\hat\beta$$

$$\leq \quad '' \quad + \quad '' \quad + \quad '' \quad + \quad '' \quad + \quad '' \quad + \lambda_3\cdot\tau_L\hat\beta^T\hat\beta$$

$$\leq L(\bar\beta|D)^{(u)}+(\beta^{(u)}-\bar\beta^{(u)})^T\nabla L(\bar\beta|D)^{(u)}+\frac{\gamma_u}{2}(\beta-\bar\beta)^T(\beta-\bar\beta)+\lambda_1\|\beta^{(u)}\|_1$$

$$+\lambda_2\cdot\|\beta^{(u)}\|_2$$

$$+\lambda_3$$

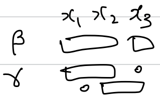
Logistic Network.

$-\nabla L(\beta \mid D)$	H
$\frac{1}{n}\sum_{i=1}^{n} \tau_i (y_i - \underline{x}_i^T \underline{\beta}) \cdot \underline{x}_i$	$\frac{1}{n} X^T \Gamma X$
$\frac{1}{n}\sum_{i=1}^{n} \tau_i \cdot y_i \cdot \underline{x}_i \cdot \frac{1}{1+\exp(y_i \underline{x}_i^T \underline{\beta})}$	$\frac{1}{4n} X^T \Gamma X$

Overlapping Group Lasso.

Model: $y_i = \beta_0 + x_i^T \beta + \varepsilon_i$.

$$= \beta_0 + \sum_{g=1}^{G} \sum_{j \in A_g} x_{ij} \cdot \gamma_j \qquad \sum_{g} \sum_{j \in A_g} \gamma_j = \sum_{g} \sum_{j \in A_g} \beta_j$$



$$\beta = \sum \gamma_{(j)}, \quad \beta, \gamma_{(j)} \in \mathbb{R}^p$$

Let $\Gamma = [\gamma_{(1)}, \gamma_{(2)}, \cdots, \gamma_{(J)}] \in \mathbb{R}^{p \times J}$

$\|\gamma_{(1)}\|, \|\gamma_{(2)}\|, \cdots, \|\gamma_{(J)}\|$

Problem:

Overlapping Group Lasso: $\underset{\{\gamma_{(j)}\}_{j=1,\cdots,J}}{\arg\max} \ \mathcal{L}(\gamma | D) + \lambda \cdot \sum_{j=1}^{J} \sqrt{d_j} \cdot \|\gamma_{(j)}\|_2$

Group Lasso Net: $\underset{\beta}{\arg\max} \ \mathcal{L}(\beta | D) + \lambda_1 \cdot \sum_{k=1}^{K} \sqrt{d_k} \|\beta_k\|_2 + \lambda_2 \cdot \beta^T L \beta$

Overlapping Group Lasso Net: $\arg\max \ \mathcal{L}(\gamma | D) + \lambda_1 \cdot \sum\sum \sqrt{d_k} \|\gamma_{(j)}\|_2 + \lambda_2 \cdot \sum_{u \sim v} \left\{ \frac{\|\gamma_{(u)}\|_2}{\sqrt{d_u}} - \frac{\|\gamma_{(v)}\|_2}{\sqrt{d_u}} \right\}^2$