13. elastic net solves:

$$\min_{\vec{w}\in\mathbb{R}^{d}} \frac{1}{N}\|\vec{y}-X\vec{w}\|_2^2 + \frac{\lambda_1}{N}\underbrace{\|\vec{w}\|_1}_{\text{not smooth}} + \frac{\lambda_2}{N}\|\vec{w}\|_2^2$$

$\therefore$ not smooth

$\therefore$ use coordinate descent (instead of gradient descent):

$$w_i^{(t+1)} = \underset{w_i\in\mathbb{R}}{\operatorname{argmin}} \frac{1}{N}\sum_{n=1}^{N}\left(y_n - \sum_{j\neq i} w_j^{(t)}x_{n,j} - w_i x_{n,i}\right)^2 + \frac{\lambda_1}{N}\left(\underbrace{\sum_{j\neq i}|w_j^{(t)}|}_{\text{constant}} + |w_i|\right) + \frac{\lambda_2}{N}\left(\underbrace{\sum_{j\neq i}(w_j^{(t)})^2}_{\text{constant}} + w_i^2\right)$$

$\rightarrow$ closed form sol: $w_i^{(t+1)} \leftarrow \underset{[-1,1]}{\alpha}\cdot \max(\underset{\mathbb{R}}{\beta}, 0)$ (sign flip)

if $\beta \le 0$, $w_i^{(t+1)} = 0$ $\rightarrow$ sparsity

$$\left(y_n - \sum_{j\neq i} w_j^{(t)}x_{n,j} - w_i x_{n,i}\right)^2$$
$$= \left[\left(y_n - \sum_{j\neq i} w_j^{(t)}x_{n,j}\right) - w_i x_{n,i}\right]^2$$
$$= \left(y_n - \sum_{j\neq i} w_j^{(t)}x_{n,j}\right)^2 - 2\left(y_n - \sum_{j\neq i} w_j^{(t)}x_{n,j}\right)w_i x_{n,i} + (w_i x_{n,i})^2$$
$$= y_n^2 - 2y_n\sum_{j\neq i} w_j^{(t)}x_{n,j} + \left(\sum_{j\neq i} w_j^{(t)}x_{n,j}\right)^2 - 2y_n w_i x_{n,i} + 2\sum_{j\neq i} w_j^{(t)}x_{n,j}w_i x_{n,i} + (w_i x_{n,i})^2$$

expand the update equation:

$$w_i^{(t+1)} = \underset{w_i\in\mathbb{R}}{\operatorname{argmin}}\Bigg\{\frac{1}{N}\sum_{n=1}^{N}\left[y_n^2 - 2y_n\sum_{j\neq i} w_j^{(t)}x_{n,j} + \left(\sum_{j\neq i} w_j^{(t)}x_{n,j}\right)^2 - 2y_n w_i x_{n,i} + 2\sum_{j\neq i} w_j^{(t)}x_{n,j}w_i x_{n,i} + (w_i x_{n,i})^2\right]$$
$$+ \frac{\lambda_1}{N}\sum_{j\neq i}|w_j^{(t)}| + \frac{\lambda_1}{N}|w_i| + \frac{\lambda_2}{N}\sum_{j\neq i}(w_j^{(t)})^2 + \frac{\lambda_2}{N}w_i^2\Bigg\}$$

$$= \underset{w_i\in\mathbb{R}}{\operatorname{argmin}}\Bigg[\cancel{1} - \cancel{\frac{2}{N}\sum_{n=1}^{N}y_n\sum_{j\neq i}w_j^{(t)}x_{n,j}} + \frac{1}{N}\sum_{n=1}^{N}\cancel{\left(\sum_{j\neq i}w_j^{(t)}x_{n,j}\right)^2} - \frac{2w_i}{N}\sum_{n=1}^{N}x_{n,i} + \frac{2w_i}{N}\sum_{n=1}^{N}\sum_{j\neq i}w_j^{(t)}x_{n,j}x_{n,i} +$$
$$\frac{w_i^2}{N}\sum_{n=1}^{N}(x_{n,i})^2 + \underbrace{\cancel{\frac{\lambda_1}{N}\sum_{j\neq i}|w_j^{(t)}|}}_{\text{constant}} + \frac{\lambda_1}{N}|w_i| + \cancel{\frac{\lambda_2}{N}\sum_{j\neq i}(w_j^{(t)})^2} + \frac{\lambda_2}{N}w_i^2\Bigg]$$

$$= \underset{w_i\in\mathbb{R}}{\operatorname{argmin}} \frac{2w_i}{N}\sum_{n=1}^{N}\left(\sum_{j\neq i}w_j^{(t)}x_{n,j}x_{n,i} - x_{n,i}\right) + \left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right)w_i^2 + \frac{\lambda_1}{N}|w_i|$$

Cases:

If $w_i > 0$, then we get:

$$\underset{w_i\in\mathbb{R}}{\operatorname{argmin}} \frac{2w_i}{N}\sum_{n=1}^{N}\left(\sum_{j\neq i}w_j^{(t)}x_{n,j}x_{n,i} - x_{n,i}\right) + \left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right)w_i^2 + \frac{\lambda_1}{N}w_i$$

$$\Rightarrow \frac{\partial}{\partial w_i}\left(\frac{2w_i}{N}\sum_{n=1}^{N}\left(\sum_{j\neq i}w_j^{(t)}x_{n,j}x_{n,i} - x_{n,i}\right) + \left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right)w_i^2 + \frac{\lambda_1}{N}w_i\right) = 0$$

$$\to \frac{2}{N}\left[\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right] + 2w_i\left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right) + \frac{\lambda_1}{N} = 0$$

$$\to 2w_i\left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right) = -\frac{2}{N}\left[\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right] - \frac{\lambda_1}{N}$$

$$\to \boxed{w_i = \frac{-\frac{2}{N}\left[\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right] - \frac{\lambda_1}{N}}{2\left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right)}}$$

case 2:

else if $w_i < 0$, then we get:

$$\underset{w_i \in \mathbb{R}}{\operatorname{argmin}} \quad \frac{2w_i}{N}\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right) + \left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right)w_i^2 - \frac{\lambda_1}{N}w_i$$

$$\to \frac{\partial}{\partial w_i}\left(\frac{2w_i}{N}\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right) + \left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right)w_i^2 - \frac{\lambda_1}{N}w_i\right) = 0$$

$$\to \frac{2}{N}\left[\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right] + 2w_i\left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right) - \frac{\lambda_1}{N} = 0$$

$$\to \boxed{w_i = \frac{-\frac{2}{N}\left[\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right] + \frac{\lambda_1}{N}}{2\left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right)}}$$

$\therefore$ For case 1:

$$-\frac{2}{N}\left[\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right] - \frac{\lambda_1}{N} > 0 \quad (\because \text{this case is } w_i > 0)$$

$$\to -\frac{2}{N}\left[\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right] > \frac{\lambda_1}{N} > 0$$

$$\to \underline{\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right) < -\frac{\lambda_1}{2} < 0} \quad \triangle$$

For case 2:

$$-\frac{2}{N}\left[\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right] + \frac{\lambda_1}{N} < 0$$ (if this case is $w_i < 0$)

$$\rightarrow -\frac{2}{N}\left[\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right] > -\frac{\lambda_1}{N} < 0$$

$$\rightarrow \sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right) > \frac{\lambda_1}{2} > 0$$

The two results shows that, if

$$\left|\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right| > \frac{\lambda_1}{2},$$

then we should update w_i as

$$\operatorname{sign}\left(\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right) \frac{\left|-\frac{2}{N}\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right| - \frac{\lambda_1}{N}}{2\left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right)}$$

Therefore, we have $\alpha = \operatorname{sign}\left(\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right)$

$$\beta = \frac{\left|-\frac{2}{N}\sum_{n=1}^{N}\left(\sum_{j\neq i} w_j^{(t)} x_{n,j} x_{n,i} - x_{n,i}\right)\right| - \frac{\lambda_1}{N}}{2\left(\frac{\sum_{n=1}^{N}(x_{n,i})^2}{N} + \frac{\lambda_2}{N}\right)}$$

$\square$