Optimization and Machine Learning
Assignment 1
Due Date: 22 March 2023, 5 pm

Question 1. We show the steps to obtain the given analytical solution of the one-dimensional subproblem for $j \neq 0$ in this assignment.

We recall the Coordinate Descent Algorithm with weighted samples first.

Let's say we have the following data $\{(x_i, y_i): i = 1, \ldots, n\}$, $x_i \in \mathbb{R}^p, y_i \in \mathbb{R}$, scaled and following input $\lambda > 0, \alpha \in [0,1]$, $\overline{\beta_0}$, $\bar{\beta} = \text{argmin}_{\beta\beta_0} \frac{1}{2} \sum_{i=1}^{n} w_i \, (y_i - \beta_0 - x_i^T \beta)^2 + \frac{\lambda}{2} ||\beta||_2^2$ /*Analytical Solution (from the slides)*/.

In the coordinate descent with weighted samples, $\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_{ij}} = 0$, $\frac{1}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_{ij}^2} = 1$, $j = 1, 2, \ldots, p$.

We aim to minimize the following term:

$$\min_{\beta_p \, (i=0,1,\ldots,p)} \frac{1}{2} \sum_{i=1}^{n} w_i \, (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \sum_{i=1}^{p} (\alpha {\beta_i}^2 + (1-\alpha)|\beta_j|).$$

∴It is to minimize the following term:

$$\min_{\beta_j} \frac{1}{2} \sum_{i=1}^{n} w_i (y_i - \overline{\beta_0} - \sum_{k \neq j} x_{ij} \overline{\beta_k} - x_{ij} \beta_i)^2 + \lambda (\frac{1-\alpha}{2} \sum_{k \neq j} \widehat{\beta_J} + \frac{1-\alpha}{2} \widehat{\beta_J^2} + \alpha \sum_{k \neq j} |\widehat{\beta_k}| + \alpha |\beta_j|)$$ where $\overline{\beta_J}$ is defined as:

$$\therefore \overline{\beta_J} = \text{argmin} \frac{1}{2n} \sum_{i=1}^{n} w_i (\overline{y_i} - x_{ij} \beta_i)^2 + \lambda \frac{1-\alpha}{2} {\beta_j}^2 + \lambda \alpha |\beta_j|$$

$$= \underset{\beta_j}{\text{argmin}} \ \frac{1}{2n} \sum_{i=1}^{n} w_i \big(y_i - x_{ij} \beta_i\big)^2 + \lambda \frac{1-\alpha}{2} {\beta_j}^2 + \lambda \alpha |\beta_j|$$

$$= \underset{\beta_j}{\text{argmin}} \ \frac{1}{2n} w_i (\sum_{i=1}^{n} |y_i|^2 - \sum_{i=1}^{n} 2 \, x_{ij} \overline{y_i} \beta_i + \sum_{i=1}^{n} x_{ij}^2 {\beta_j}^2) + \lambda \frac{1-\alpha}{2} {\beta_j}^2 + \lambda \alpha |\beta_j|$$

Let's define $\widehat{\beta_J}$ as follows:

$$\because \widehat{\beta_J} = \underset{\beta_j}{\text{argmin}} \frac{1}{2} \sum_{i=1}^{n} w_i \big(y_i - x_{ij} \beta_i\big)^2 \Rightarrow \widehat{\beta_J} = \frac{1}{n} \sum_{i=1}^{n} w_i x_{ij} \, \overline{y_i}$$

If $j \neq 0$, $\widehat{\beta_J} = \sum_{i=1}^{n} w_i \, x_{ij} (y_i - \overline{\beta_0} - \sum_{k \neq j} x_{ik} \, \overline{\beta_k})$.

So, when we plug $\widehat{\beta_J}$ into $\overline{\beta_J}$, we get:

$$\therefore \overline{\beta_J} = \underset{\beta_j}{\operatorname{argmin}} \frac{1}{2n}\sum_{i=1}^{n} w_i \left(\widehat{\beta_J}^2 - 2\widehat{\beta_J}\beta_j + {\beta_j}^2\right) + \lambda\frac{1-\alpha}{2}{\beta_j}^2 + \lambda\alpha|\beta_j|$$

$$= \underset{\beta_j}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^{n} w_i \left(\beta_j - \widehat{\beta_J}\right)^2 + \lambda\frac{1-\alpha}{2}{\beta_j}^2 + \lambda\alpha|\beta_j|$$

$$\therefore \overline{\beta_J} = \begin{cases} \dfrac{\widehat{\beta_J} - \lambda\alpha}{\sum_{i=1}^{n} w_i\, x_{ij}^2 + \lambda(1-\alpha)}, \text{if } \widehat{\beta_J} > 0 \text{ and } \lambda\alpha < \widehat{\beta_J} \\ \dfrac{\widehat{\beta_J} + \lambda\alpha}{\sum_{i=1}^{n} w_i\, x_{ij}^2 + \lambda(1-\alpha)}, \text{if } \widehat{\beta_J} < 0 \text{ and } \lambda\alpha < -\widehat{\beta_J} \\ 0, \text{otherwise} \end{cases}$$

End of steps.