1 L0.5 for Section 3.5

For the penalty term in the optimization problem, we choose q = 0.5 i.e., 0<q<1

$\hat{\theta}_j = \text{argmin}_\theta(\frac{1}{2}\sum_{i=1}^{N}[Y_i - \sum_{k=1}^{p} X_{i_k}\theta_k]^2 + \lambda\sum_{k=1}^{p}\sqrt{|\theta_k|}])$

The math is very similar to the one for LASSO as shown above(Section 3.2). So, we have left the details to the appendix.

$\theta_j = \frac{sign(T).(|T| - \frac{\lambda}{2\sqrt{\theta_j}})}{N-1}$

Where, A = N-1 ; B = $\frac{\lambda}{2\sqrt{\theta_j}}$

As seen here, when 0<q<1, the solution is in terms of itself. The solution here is non-trivial or non-convex. This solution is in our desired form of Pathwise Coordinate Descent. We have left the implementation as a future scope of the project.

2 L0.5 for Appendix

For the penalty term in the optimization problem, we choose q = 0.5 i.e., 0<q<1

$\hat{\theta}_j = \text{argmin}_\theta(\frac{1}{2}\sum_{i=1}^{N}[Y_i - \sum_{k=1}^{p} X_{i_k}\theta_k]^2 + \lambda\sum_{k=1}^{p}\sqrt{|\theta_k|}])$

We consider the tuning parameter $\lambda \geq 0$. Now, we consider the above optimization problem as 'O' and take its derivative with respect to θ_j to find the stationary points.

$\frac{\partial O}{\partial \theta_j} = -\sum_{i=1}^{N} x_{ij}(Y_i - x_{ij}\theta_j - \sum_{k\neq j}^{p} x_{ij}\theta_j) + \frac{\lambda sign(\theta_j)}{2\sqrt{|\theta_j|}}$

$\frac{\partial O}{\partial \theta_j} = \sum_{i=1}^{N} x_{ij}(Y_i - \sum_{k\neq j}^{p} x_{ij}\theta_j) - \sum_{i+1}^{N} {x_{ij}}^2\theta_j + \frac{\lambda sign(\theta_j)}{2\sqrt{|\theta_j|}}$

Now, since our data is standardised,

$\sum_{i+1}^{N} {x_{ij}}^2 =$ N - 1

Setting the derivative to 0.
$\sum_{i=1}^{N} x_{ij}(Y_i - x_{ij}\theta_j) = (N-1)\theta_j + \frac{\lambda sign(\theta_j)}{2\sqrt{|\theta_j|}}$ T = $\sum_{i=1}^{N} x_{ij}(Y_i - x_{ij}\theta_j)$

As described before for LASSO, we can divide this into a 2-case argument and parameterize it in terms of A and B for implementing PCD.
The 2-case argument is:

1. T $\geq \frac{\lambda}{2\sqrt{|\theta_j|}}$

2. $\theta_j \geq 0$ if T ≥ 0 else $\theta_j < 0$ if T < 0

Finally, we can write our solution as:

$\theta_j = \frac{sign(T)(|T| - \frac{\lambda}{2\sqrt{|\theta_j|}})}{N-1}$

Where,

A = N-1 ; B = $\frac{\lambda}{2\sqrt{\theta_j}}$