Proposed Formulation for feature selection

L0 Norm with Group Constraints:

The objective function consists of the sum of squared errors for the regression model, a penalty term proportional to the L0 norm of the coefficients (modeled by the binary variables), and a group sparsity penalty term. The constraints ensure that coefficients are either fixed at zero or allowed to vary freely when their corresponding binary indicators are active, and they maintain the logical connections between individual coefficients and their respective groups.

$$\min_{\gamma,\gamma_{\text{nonzero}},\text{group}} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{p} \sum_{k=1}^{r} Z_{ijk}\gamma_{jk} \right)^2 + \lambda \sum_{j=1}^{p} \sum_{k=1}^{r} \gamma_{\text{nonzero},jk} + \lambda_{\text{group}} \sum_{j=1}^{p} \text{group}_j$$

Subject to the following constraints:

Coefficient-Binary Link Constraints:

$$-\text{BIG\_M} \cdot \gamma_{\text{nonzero},jk} \leq \gamma_{jk} \leq \text{BIG\_M} \cdot \gamma_{\text{nonzero},jk} \quad \forall j = 1, ..., p; \; k = 1, ..., r$$

These constraints ensure that the regression coefficients γ_{jk} can only be non-zero when their corresponding binary variables $\gamma_{\text{nonzero},jk}$ are set to 1.

The use of a large value BIG_M effectively relaxes the constraint when $\gamma_{\text{nonzero},jk}$ is 1.

Group Limit Constraint:

$$\sum_{j=1}^{p} \text{group}_j \leq \text{group\_limit}$$

This constraint enforces that the total number of groups with non-zero coefficients does not exceed a predefined limit group_limit, thus controlling the group-level sparsity.

Group-Nonzero Coefficient Link Constraints:

$$\sum_{k=1}^{r} \gamma_{\text{nonzero},jk} \leq r \cdot \text{group}_j \quad \forall j = 1, ..., p$$

These constraints ensure that if any coefficient in a group is non-zero, the group binary variable group_j will be set to 1. They link the individual coefficient non-zero indicators to the group indicator.

Binary Coefficient Constraint within Group:

$$\gamma_{\text{nonzero},jk} \leq \text{group}_j \quad \forall j = 1, ..., p; \ k = 1, ..., r$$

These constraints enforce that individual coefficients can only be non-zero if the entire group is selected, meaning that group_j must be 1 if any $\gamma_{\text{nonzero},jk}$ is 1.

And the decision variables are bounded as:

$$\gamma_{jk} \in \mathbb{R}, \ \ \gamma_{\text{nonzero},jk} \in \{0, 1\} \quad \forall j = 1, ..., p; \ k = 1, ..., r$$

$$\text{group}_j \in \{0, 1\} \quad \forall j = 1, ..., p$$

These constraints define the domains of the decision variables, stating that regression coefficients are real numbers, whereas the binary variables indicating non-zero coefficients and group selections are binary.

In this formulation:

- γ_{jk} are the continuous decision variables representing the coefficients of the regression model.
- $\gamma_{\text{nonzero},jk}$ are binary variables indicating whether the corresponding coefficient γ_{jk} is non-zero.
- group_j are binary variables indicating whether any coefficients in the j-th group are non-zero.
- λ and λ_{group} are regularization parameters controlling the sparsity of coefficients and the sparsity of group selection, respectively.
- BIG_M is a sufficiently large positive constant used in the "big M" method to effectively turn constraints on or off based on the value of the binary variables.
- group_limit represents the maximum number of groups that can have non-zero coefficients.