

# Proposed Formulation for Feature Selection

## L1 Norm with Group Constraints:

The following model incorporates both L1 regularization, which induces sparsity in the regression coefficients (  $\gamma$  ), and group selection, which allows for simultaneous selection or deselection of entire groups of predictors.

Minimize the objective function:

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

Subject to:

### 1. Coefficient Magnitude Constraints:

$$\gamma_{jk} \leq t_{jk}, \quad -\gamma_{jk} \leq t_{jk}, \quad \forall j \in \{1, \dots, p\}, \forall k \in \{1, \dots, r\}$$

These constraints ensure that the auxiliary variable  $t_{jk}$  is at least the absolute value of the regression coefficient  $\gamma_{jk}$ , effectively allowing  $t_{jk}$  to capture the L1 norm of  $\gamma_{jk}$ .

### 2. Group Constraints for Coefficients:

$$\gamma_{jk} \leq \text{BIG\_M} \times \gamma_{\text{nonzero}_{jk}}, \quad -\gamma_{jk} \geq -\text{BIG\_M} \times \gamma_{\text{nonzero}_{jk}}, \quad \forall j \in \{1, \dots, p\}, \forall k \in \{1, \dots, r\}$$

$$\gamma_{\text{nonzero}_{jk}} \leq \text{group}_j, \quad \forall j \in \{1, \dots, p\}, \forall k \in \{1, \dots, r\}$$

These constraints ensure that if a group variable  $\text{group}_j$  is zero, then all corresponding regression coefficients  $\gamma_{jk}$  for that group must be zero. If  $\text{group}_j$  is one, then  $\gamma_{jk}$  can take any value within the bounds set by  $\text{BIG\_M}$ , the big M constant.

### 3. Group Selection to Coefficient Nonzero Constraint:

$$\sum_{k=1}^r \gamma_{\text{nonzero}_{jk}} \geq \text{group}_j, \quad \forall j \in \{1, \dots, p\}$$

This constraint ensures that if any  $\gamma_{jk}$  for a group is nonzero, the group must be selected, i.e.,  $\text{group}_j$  must be one. Conversely, if all  $\gamma_{jk}$  for a group are zero,  $\text{group}_j$  can be zero, indicating deselection of the group.

### 4. Group Limit Constraint (Optional):

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

This constraint limits the number of groups that can be selected, enforcing sparsity at the group level. The `group_limit` is a parameter that can be defined by the user.

#### 5. Binary and Non-Negativity Constraints:

$$\text{group}_j \in \{0, 1\}, \quad t_{jk} \geq 0, \quad \gamma_{\text{nonzero}_{jk}} \in \{0, 1\}, \quad \forall j \in \{1, \dots, p\}, \forall k \in \{1, \dots, r\}$$

$\text{group}_j$  and  $\gamma_{\text{nonzero}_{jk}}$  are binary variables that indicate the inclusion of the  $j$ -th group of predictors in the model and whether the coefficient  $\gamma_{jk}$  is nonzero, respectively. The  $t_{jk}$  variables are constrained to be non-negative as they represent the magnitude of the coefficients.

Where:

- $Y_i$ : The observed response for the  $i$ -th observation.
- $Z_{ijk}$ : The value of the  $k$ -th feature for the  $i$ -th observation in the  $j$ -th group.
- $\gamma_{jk}$ : The coefficient of the  $k$ -th feature for the  $j$ -th group.
- $t_{jk}$ : The auxiliary variable used to linearize the L1 norm of the coefficient  $\gamma_{jk}$ .
- $\text{group}_j$ : The binary variable indicating whether the  $j$ -th group of features is selected.
- $\gamma_{\text{nonzero}_{jk}}$ : The auxiliary binary variable indicating whether  $\gamma_{jk}$  is nonzero.
- $\text{BIG\_M}$ : A large positive constant used in the big M method.
- $\lambda$ : The regularization parameter for the L1 norm of the coefficients.
- $\lambda_{\text{group}}$ : The regularization parameter for the group selection.
- `group_limit`: A user-defined limit on the number of groups that can be selected.

The objective function consists of the sum of squared errors, the L1 regularization term that promotes sparsity in the coefficients, and a group penalty term that encourages the selection or deselection of entire groups of predictors. The constraints ensure that the L1 norm is modeled correctly, the group structure is respected, and the model adheres to the predefined sparsity level.