*Last Updated: 06/11/2021*

# General MIP Original Problem for $L \geq 2$

### *with artificial variable* $g_{n,k',k,l}$

**Parameters:**

$x_{n,d}$: binary vector inputs of size n × d, where n is the number of data points, d is the number of dimensions/features

$y_n$: binary vector labeled outputs of size n × 1, where n is the number of data points

**Decision Variables:**

$\alpha_{d,k,0}$: Weight for feature d in unit k in the first hidden layer, $\forall\ d \in D,\ k \in K$

$\alpha_{k',k,l}$: Weight from the $k'^{th}$ unit in layer l-1 to unit k in layer l, $\forall\ k',\ k \in K,\ l \in \{1, 2, 3, \ldots, L-1\}$

$\alpha_{k',0,L}$: Weight from the $k'^{th}$ unit in hiddenlayer L to unit output layer L+1, $\forall\ \ k' \in K$

$\beta_{k,l}$: Bias for unit k in layer l, $\forall\ k \in K,\ l \in \{0, 1, \ldots, L-1\}$

$\beta_{0,L}$: Bias in the final layer

$h_{n,k,l}$: Binary output of unit k in layer l, $\forall\ n \in N,\ k \in K,\ l \in \{0, 1, \ldots L-1\}$

$g_{n,k',k,l}$: Artificial Variable for decomposition, $\forall\ n \in N,\ k, k' \in K,\ l \in \{1, \ldots, L-1\}$

$g_{n,k',0,L}$: Artificial Variable for decomposition of the output layer, $\forall\ n \in N,\ k' \in K$

$z_{n,k',k,l}$: Auxilliary variable that represents $\alpha_{k',k,l}h_{n,k,(l-1)}$ $\forall\ n \in N,\ k',\ k \in K,\ l \in \{1, 2, \ldots, L-1\}$

$z_{n,k',0,L}$: Auxilliary variable that represents $\alpha_{k',0,L}h_{n,k,(L-1)}$ $\forall\ n \in N,\ k',\ k \in K$

$\hat{y}_n$: Output of final layer, $\forall\ n \in N$

$\ell_n$: Absolute Misclassification of data point n, $\forall\, n \in N$

**Objective:**

$$\min_{\alpha,\beta,h,g,z,\hat{y},\ell} \sum_{n=1}^{N} \ell_n$$

**Constraints:**

$$\text{subject to} \quad \sum_{k'=0}^{K} (z_{n,k',0,L}) + \beta_{0,L} \leq -\epsilon + (M+\epsilon)\hat{y}_n, \ \forall\, n$$

$$\sum_{k'=0}^{K} (z_{n,k',0,L}) + \beta_{0,L} \geq \epsilon + (m-\epsilon)(1-\hat{y}_n), \ \forall\, n$$

$$\ell_n \geq y_n - \hat{y}_n, \ \forall\, n$$

$$\ell_n \geq -y_n + \hat{y}_n, \ \forall\, n$$

$$\sum_{d=0}^{D} (\alpha_{d,k,0} x_{n,d}) + \beta_{k,0} \leq -\epsilon + (M+\epsilon) h_{n,k,0}, \ \forall\, n, k$$

$$\sum_{d=0}^{D} (\alpha_{d,k,0} x_{n,d}) + \beta_{k,0} \geq \epsilon + (m-\epsilon)(1-h_{n,k,0}), \ \forall\, n, k$$

$$\sum_{k'=0}^{K} (z_{n,k',k,l}) + \beta_{k,l} \leq -\epsilon + (M+\epsilon) h_{n,k,l}, \ \forall\, n, k, l \in \{1, 2, \ldots, L-1\}$$

$$\sum_{k'=0}^{K} (z_{n,k',k,l}) + \beta_{k,l} \geq \epsilon + (m-\epsilon)(1-h_{n,k,l}), \ \forall\, n, k, l \in \{1, 2, \ldots, L-1\}$$

$$z_{n,k',k,l} \leq \alpha_{k',k,l}, \ \forall\, n, k', k, l \in \{1, 2, \ldots, L-1\}$$

$$z_{n,k',k,l} - \alpha_{k',k,l} \geq m(1-g_{n,k',k,l}), \ \forall\, n, k', k, l \in \{1, 2, \ldots, L-1\}$$

$z_{n,k',k,l} \leq Mg_{n,k',k,l}, \ \forall \ n, k', k, l \in \ \{1, 2, ..., L-1\}$

$z_{n,k',k,l} \geq mg_{n,k',k,l}, \ \forall \ n, k', k, l \in \ \{1, 2, ..., L-1\}$

$g_{n,k',k,l} = h_{n,k',(l-1)}, \ \forall \ n, k', k, l \in \{1, 2, ..., L-1\}$

$z_{n,k',0,L} \leq \alpha_{k',0,L}, \ \forall \ n, k'$

$z_{n,k',0,L} - \alpha_{k',0,L} \geq m(1 - g_{n,k',0,L}), \ \forall \ n, k'$

$z_{n,k',0,L} \leq Mg_{n,k',0,L}, \ \forall \ n, k'$

$z_{n,k',0,L} \geq mg_{n,k',0,L}, \ \forall \ n, k'$

$g_{n,k',0,L} = h_{n,k',(L-1)}, \ \forall \ n, k', k$

$Lower\ Bound \leq \alpha_{d,k,0}, \ \alpha_{k',k,l}, \ \alpha_{k',0,L-1} \leq Upper\ Bound, \ \forall \ d, k', k, l \in \ \{1, 2, ..., L-1\}$

$Lower\ Bound \leq \ z_{n,k',k,l}, \ z_{n,k',0,L} \leq Upper\ Bound, \ \forall \ n, k', k, l \in \ \{1, 2, ..., L-1\}$

$Lower\ Bound \leq \beta_{k,l}, \ \beta_{L-1} \leq Upper\ Bound, \ \forall \ k, l \in \ \{0, 1, 2, ..., L-1\}$

$\hat{y}_n, h_{n,k,l} \in \{0, 1\}, \ \forall \ n, k', k, l \in \ \{0, 1, ..., L-1\}$

$g_{n,k',k,l}, g_{n,k',0,L} \in \{0, 1\}, \ \forall \ n, k', k, l \in \ \{1, 2, ..., L-1\}$

$0 \leq \ell_n \leq 1, \ \forall \ n \in N$

---

## Lagrangian Relaxation and Dual

$$\zeta_{LR}(\lambda) = \min_{\alpha,\beta,h,g,z,\hat{y},\ell} \sum_{n=1}^{N} (\ell_n + \sum_{k'=0}^{K} (\sum_{k=0}^{K} (\sum_{l=1}^{L-1} (\lambda_{n,k',k,l}(g_{n,k',k,l} - h_{n,k',(l-1)}))) + \lambda_{n,k',0,L}(g_{n,k',0,L} - h_{n,k',(L-1)})))$$

subject to $$\sum_{k'=0}^{K}(z_{n,k',0,L})+\beta_{0,L}\leq -\epsilon+(M+\epsilon)\hat{y}_n,\ \forall\, n$$

$$\sum_{k'=0}^{K}(z_{n,k',0,L})+\beta_{0,L}\geq \epsilon+(m-\epsilon)(1-\hat{y}_n),\ \forall\, n$$

$$\ell_n\geq y_n-\hat{y}_n,\ \forall\, n$$

$$\ell_n\geq -y_n+\hat{y}_n,\ \forall\, n$$

$$\sum_{d=0}^{D}(\alpha_{d,k,0}x_{n,d})+\beta_{k,0}\leq -\epsilon+(M+\epsilon)h_{n,k,0},\ \forall\, n,k$$

$$\sum_{d=0}^{D}(\alpha_{d,k,0}x_{n,d})+\beta_{k,0}\geq \epsilon+(m-\epsilon)(1-h_{n,k,0}),\ \forall\, n,k$$

$$\sum_{k'=0}^{K}(z_{n,k',k,l})+\beta_{k,l}\leq -\epsilon+(M+\epsilon)h_{n,k,l},\ \forall\, n,k,l\in\ \{1,2,\ldots,L-1\}$$

$$\sum_{k'=0}^{K}(z_{n,k',k,l})+\beta_{k,l}\geq \epsilon+(m-\epsilon)(1-h_{n,k,l}),\ \forall\, n,k,l\in\ \{1,2,\ldots,L-1\}$$

$$z_{n,k',k,l}\leq \alpha_{k',k,l},\ \forall\, n,k',k,l\in\ \{1,2,\ldots,L-1\}$$

$$z_{n,k',k,l}-\alpha_{k',k,l}\geq m(1-g_{n,k',k,l}),\ \forall\, n,k',k,l\in\ \{1,2,\ldots,L-1\}$$

$$z_{n,k',k,l}\leq Mg_{n,k',k,l},\ \forall\, n,k',k,l\in\ \{1,2,\ldots,L-1\}$$

$$z_{n,k',k,l}\geq mg_{n,k',k,l},\ \forall\, n,k',k,l\in\ \{1,2,\ldots,L-1\}$$

$$z_{n,k',0,L}\leq \alpha_{k',0,L},\ \forall\, n,k'$$

$$z_{n,k',0,L}-\alpha_{k',0,L}\geq m(1-g_{n,k',0,L}),\ \forall\, n,k'$$

$$z_{n,k',0,L}\leq Mg_{n,k',0,L},\ \forall\, n,k'$$

$$z_{n,k',0,L} \geq m g_{n,k',0,L}, \ \forall \ n, k'$$

$$Lower\ Bound \leq \alpha_{d,k,0},\ \alpha_{k',k,l},\ \alpha_{k',0,L-1} \leq Upper\ Bound,\ \forall \ d, k', k, l \in \ \{1, 2, \ldots, L-1\}$$

$$Lower\ Bound \leq \ z_{n,k',k,l},\ z_{n,k',0,L} \leq Upper\ Bound,\ \forall \ n, k', k, l \in \ \{1, 2, \ldots, L-1\}$$

$$Lower\ Bound \leq \beta_{k,l},\ \beta_{L-1} \leq Upper\ Bound,\ \forall \ k, l \in \ \{0, 1, 2, \ldots, L-1\}$$

$$\hat{y}_n, h_{n,k,l} \in \{0, 1\},\ \forall \ n, k', k, l \in \ \{0, 1, \ldots, L-1\}$$

$$g_{n,k',k,l}, g_{n,k',0,L} \in \{0, 1\},\ \forall \ n, k', k, l \in \ \{1, 2, \ldots, L-1\}$$

$$0 \leq \ell_n \leq 1,\ \forall \ n \in N$$

---

## Sub-problems

$$\zeta_0(\lambda) = \min_{\alpha,\beta,h} \sum_{n=0}^{N} (\sum_{k'=0}^{K} (\sum_{k=0}^{K} (-\lambda_{n,k',k,1} h_{n,k',0})))$$

$$\text{subject to} \quad \sum_{d=0}^{D} (\alpha_{d,k,0} x_{n,d}) + \beta_{k,0} \leq \ -\epsilon + (M+\epsilon) h_{n,k,0},\ \forall \ n, k$$

$$\sum_{d=0}^{D} (\alpha_{d,k,0} x_{n,d}) + \beta_{k,0} \geq \epsilon + (m-\epsilon)(1-h_{n,k,0}),\ \forall \ n, k$$

$$Lower\ Bound \leq \alpha_{d,k,0} \leq Upper\ Bound,\ \forall \ d, k$$

$$Lower\ Bound \leq \beta_{k,0} \leq Upper\ Bound,\ \forall \ k$$

$$h_{n,k,0} \in \{0, 1\},\ \forall \ n, k$$

$$\zeta_l(\lambda) = \min_{\alpha,\beta,z,h,g} \sum_{n=0}^{N} \left(\sum_{k'=0}^{K} \left(\sum_{k=0}^{K} (\lambda_{n,k',k,l} g_{n,k',k,l} - \lambda_{n,k',k,l+1} h_{n,k',l})\right)\right)$$

subject to $\sum_{k'=0}^{K} (z_{n,k',k,l}) + \beta_{k,l} \leq -\epsilon + (M+\epsilon) h_{n,k,l}, \; \forall\, n, k$

$$\sum_{k'=0}^{K} (z_{n,k',k,l}) + \beta_{k,l} \geq \epsilon + (m-\epsilon)(1-h_{n,k,l}), \; \forall\, n, k$$

$$z_{n,k',k,l} \leq \alpha_{k',k,l}, \; \forall\, n, k', k$$

$$z_{n,k',k,l} - \alpha_{k',k,l} \geq m(1-g_{n,k',k,l}), \; \forall\, n, k', k$$

$$z_{n,k',k,l} \leq M g_{n,k',k,l}, \; \forall\, n, k', k$$

$$z_{n,k',k,l} \geq m g_{n,k',k,l}, \; \forall\, n, k', k$$

$$\textit{Lower Bound} \leq \alpha_{k',k,l} \leq \textit{Upper Bound}, \; \forall\, k', k$$

$$\textit{Lower Bound} \leq z_{n,k',k,l} \leq \textit{Upper Bound}, \; \forall\, n, k, k'$$

$$\textit{Lower Bound} \leq \beta_{k,l} \leq \textit{Upper Bound}, \; \forall\, k$$

$$h_{n,k,l}, g_{n,k',k,l} \in \{0, 1\}, \; \forall\, n, k, k'$$

$$\zeta_{L-1}(\lambda) = \min_{\alpha,\beta,z,h,g} \sum_{n=0}^{N} \left(\sum_{k'=0}^{K} \left(\sum_{k=0}^{K} (\lambda_{n,k',k,L-1} g_{n,k',k,L-1}) - \lambda_{n,k',0,L} h_{n,k',L-1}\right)\right)$$

subject to $\sum_{k'=0}^{K} (z_{n,k',k,L-1}) + \beta_{k,L-1} \leq -\epsilon + (M+\epsilon) h_{n,k,L-1}, \; \forall\, n, k$

$$\sum_{k'=0}^{K} (z_{n,k',k,L-1}) + \beta_{k,L-1} \geq \epsilon + (m-\epsilon)(1-h_{n,k,L-1}), \; \forall\, n, k$$

$$z_{n,k',k,L-1} \leq \alpha_{k',k,L-1}, \; \forall\, n, k', k$$

$$z_{n,k',k,L-1} - \alpha_{k',k,L-1} \geq m(1 - g_{n,k',k,L-1}),\ \forall\, n, k', k$$

$$z_{n,k',k,L-1} \leq Mg_{n,k',k,L-1},\ \forall\, n, k', k$$

$$z_{n,k',k,L-1} \geq mg_{n,k',k,L-1},\ \forall\, n, k', k$$

$$Lower\ Bound \leq \alpha_{k',k,L-1} \leq Upper\ Bound,\ \forall\, k', k$$

$$Lower\ Bound \leq z_{n,k',k,L-1} \leq Upper\ Bound,\ \forall\, n, k, k'$$

$$Lower\ Bound \leq \beta_{k,L-1} \leq Upper\ Bound,\ \forall\, k$$

$$h_{n,k,L-1}, g_{n,k',k,L-1} \in \{0, 1\},\ \forall\, n, k, k'$$

$$\zeta_L(\lambda) = \min_{\alpha,\beta,z,g,\hat{y},\ell} \sum_{n=0}^{N} (\ell_n + \sum_{k'=0}^{K} (\lambda_{n,k',0,L} g_{n,k',0,L}))$$

$$\text{subject to} \quad \sum_{k'=0}^{K} (z_{n,k',0,L}) + \beta_{0,L} \leq -\epsilon + (M + \epsilon)\hat{y}_n,\ \forall\, n$$

$$\sum_{k'=0}^{K} (z_{n,k',0,L}) + \beta_{0,L} \geq \epsilon + (m - \epsilon)(1 - \hat{y}_n),\ \forall\, n$$

$$\ell_n \geq y_n - \hat{y}_n,\ \forall\, n$$

$$\ell_n \geq -y_n + \hat{y}_n,\ \forall\, n$$

$$z_{n,k',0,L} \leq \alpha_{k',0,L},\ \forall\, n, k'$$

$$z_{n,k',0,L} - \alpha_{k',0,L} \geq m(1 - g_{n,k',0,L}),\ \forall\, n, k'$$

$$z_{n,k',0,L} \leq Mg_{n,k',0,L},\ \forall\, n, k'$$

$$z_{n,k',0,L} \geq mg_{n,k',0,L},\ \forall\, n, k'$$

$$Lower\ Bound \leq \alpha_{k',0,L} \leq Upper\ Bound,\ \forall\, k'$$

$$Lower\ Bound \leq z_{n,k',0,L} \leq Upper\ Bound,\ \ \forall\ n, k'$$

$$Lower\ Bound \leq \beta_{0,L} \leq Upper\ Bound$$

$$\hat{y}_n, g_{n,k',0,L} \in \{0, 1\},\ \ \forall\ n, k'$$

$$0 \leq \ell_n \leq 1,\ \ \forall\ n$$

$$z_{\mathrm{L}} = \max_{\lambda \in \mathrm{R}^{m_1}} \zeta_0(\lambda) + \sum_{l=1}^{L-2} (\zeta_l(\lambda)) + \zeta_{L-1}(\lambda) + \zeta_L(\lambda)$$

---

## Vrishabh's Notes

- Not sure if the $z \leq \alpha$ constraints are in the right subproblems. They're in their respective layers, but since neither $z$ nor $\alpha$ are in the dualized constraints, should they be in the final "master" dual problem?
- Not sure if it's okay to have $h_{n,k',l}$ in the objective, but $h_{n,k,l}$ in the constraints for $\zeta_0(\lambda),\ \zeta_l(\lambda),\ \text{and}\ \zeta_{L-1}(\lambda)$. The variables don't necessarily align