Last Updated: 06/10/2021

General MIP Original Problem for $L \geq 2$

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, $\forall\ k \in K,\ \ l \in \{0, 1, \ldots, L-1\}$

β_L: Bias in the final layer

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

$z_{n,k',k,l}$: 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, $\forall\ n \in N$

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

Objective:

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

Constraints:

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

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

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

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

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

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

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

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

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

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

$$mh_{n,k',(l-1)} \leq z_{n,k',k,l} \leq Mh_{n,k',(l-1)}, \ \forall\, n, k, k', l \in \{1, 2, \ldots, L-1\}$$

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

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

$$mh_{n,k',L-1} \leq z_{n,k',0,L} \leq Mh_{n,k',L-1},\ \forall\ n, k'$$

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

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

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

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

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

Lagrangian Relaxation and Dual

$$\zeta_{LR}(\lambda) = \min_{\alpha,\beta,z,h,\hat{y},\ell} \sum_{n=0}^{N} (\ell_n + (\sum_{k'=0}^{K} (\sum_{k=0}^{K} (\sum_{l=1}^{L-1} (\lambda_{2,n,k',k,l}(-z_{n,k',k,l} + \alpha_{k',k,l} + m(1 - h_{n,k',l-1})) + \lambda_{3,n,k',k,l}(mh_{n,k',l-1} - z_{n,k',k,l}) + \lambda_{4,n,k',k,l}(z_{n,k',k,l} - Mh_{n,k',}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Sub-problems

$$\zeta_0(\lambda)\ =\ \min_{\alpha,\beta,h}\sum_{n=0}^{N}(\sum_{k'=0}^{K}(\sum_{k=0}^{K}h_{n,k',0}(-m\lambda_{2,n,k',k,1}+m\lambda_{3,n,k',k,1}-M\lambda_{4,n,k',k,1})))+\mu_1\sum_{d=0}^{D}(\sum_{k=0}^{K}(|\alpha_{d,k,0}|))$$

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

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

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

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

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

$$\zeta_l(\lambda) = \min_{z,\beta,h,\alpha} \sum_{n=0}^{N} (\sum_{k'=0}^{K} (\sum_{k=0}^{K} (z_{n,k',k,l}(-\lambda_{2,n,k',k,l} - \lambda_{3,n,k',k,l} + \lambda_{4,n,k',k,l}) + h_{n,k',l}(-m\lambda_{2,n,k',k,l+1} + m\lambda_{3,n,k',k,l+1} - M\lambda_{4,n,k',k,l+1}) + \alpha_{k',k,l}(\lambda_{2,n,k',k,l}))))$$

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

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

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

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

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

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

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

$$\zeta_{L-1}(\lambda) = \min_{z,\beta,h,\alpha} \sum_{n=0}^{N} (\sum_{k'=0}^{K} (\sum_{k=0}^{K} (z_{n,k',k,L-1}(-\lambda_{2,n,k',k,L-1} - \lambda_{3,n,k',k,L-1} + \lambda_{4,n,k',k,L-1}) + \alpha_{k',k,L-1}(\lambda_{2,n,k',k,L-1})) + h_{n,k',L-1}(-m\lambda_{2,n,k',0,L} + m\lambda_{3,n,k',0,}$$

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

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

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

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

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

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

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

$$\zeta_L(\lambda) = \min_{z,\alpha,\beta,\hat{y},\ell} \sum_{n=0}^{N} (\ell_n + (\sum_{k'=0}^{K} z_{n,k',0,L}(-\lambda_{2,n,k',0,L} - \lambda_{3,n,k',0,L} + \lambda_{4,n,k',0,L}) + \alpha_{k',0,L}(\lambda_{2,n,k',0,L}))) + \mu_3 \sum_{k=0}^{K} (|\alpha_{k',0,L}|))$$

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

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

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

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

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

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

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

$$Lower\ Bound \le \beta_L \le Upper\ Bound$$

$$\hat{y}_n \in \{0,1\},\ \forall\ n$$

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

Master Problem

$$z_{\mathrm{L}} = \max_{\lambda \in \mathrm{R}_{+}^{m_1}} \sum_{l=1}^{L-2} (\zeta_l(\lambda)) + \zeta_{L-1}(\lambda) + \zeta_0(\lambda) + \zeta_L(\lambda) + \sum_{n=0}^{N} (\sum_{k'=0}^{K} (\sum_{k=0}^{K} (\sum_{l=1}^{L-1} m\lambda_{2,n,k',k,l})) + m\lambda_{2,n,k',0,L})$$

===

Notes and Questions for 05/03/2021

- Lagrange Dual
  - Clean up the formulation to merge the $z_{n,k',k,l}$ and respective $h_{n,k',l}$
  - Specifically general to have a sub problem for the first layer, middle layers and output layer separately.
- Read up on MIP and NN papers
  - Laurent El-Ghaoui (UC Berkeley)

Notes and Questions for 05/11/2021

- Lagrange Dual
  - Min in the sub problems over the decision variables
  - Decompose the first subproblem
- Read up on MIP and NN papers/textbooks
  - Specifically, subgradient descent in lagrange decomposition problems.

    Notes:
  - Subgradient vector of $\zeta_{LR}(\lambda)$, s^t, consists of the dualized constraints at λ^t
  - 1. Choose a starting point for all λ
    2. Let $s^t = b - Ax^t$ of the $\zeta(\lambda^t)$. If $s^t = 0$, stop
    3. $\lambda^{t+1} = \max\,(0, \lambda^t + \gamma^t s^t)$ where γ is the step size
    4. Increment t and go to 2.

    Referenced from: *Lagrangian Relaxation: An overview*; Discrete Mathematics for Bioinformatics WS 07/08, G. W. Klau, 18. Dezember 2007, 14:21

Lagrangian Relaxation, Knut Reinert, Gunnar Klau, 18. Dezember 2007, 14:21 4

Solving the Lagrangian Dual (5)

Held and Karp proposed the following formula for adapting the stepsize:

$$\gamma^t = \mu^t \frac{Z^* - Z(\lambda^t)}{\sum_{i=1}^{m}(b_i - \sum_{j=1}^{n} a_{ij}x^t)^2} \; ,$$

where

- Z^* is the value of the best solution for the original problem found so far
- μ^t is a decreasing adaption parameter with $0 < \mu^0 \leq 2$ and

$$\mu^{t+1} = \begin{cases} \alpha\mu^t & Z_D \text{ did not increase in the last } T \text{ iterations} \\ \mu^t & \text{otherwise} \end{cases}$$

  with parameters $0 < \alpha < 1$ and $T > 1$.
- the denominator is the square of the length of the subgradient vector $b - Ax^t$.

- *General Notes:*
  - Subgradient vs Cutting Plane? (Prof. Luedtke's Slides)
  - Bundle Method? (Prof. Luedtke's Slides)
  - Verify if our problem is convex (Since the relaxation is till a MIP, I'm guessing it's nonconvex)
  - Ballstep Subgradient Method (https://www.researchgate.net/publication/220442497_Lagrangian_Relaxation_via_Ballstep_Subgradient_Methods) "can provide a solution to the problem at no extra cost". Scope for not just giving us the bounds? Example was for "Lagrangian relaxation of convex programs"

Notes and Questions for 06/01/2021

- Global Dependence (Introduction to Linear Optimization - Bertsimas, Tsitsikilis)
- Code up the problems (Langrangian Dual Method-Implementation)