Boosting for Classification and Regression

Ai Kagawa [1], Jonathan Eckstein [1,2] Noam Goldberg [3]

[1]RUTCOR, Rutgers University
[2]Department of Management Science and Information Systems, Rutgers University
[3]Department of Management, Bar-Ilan University

9/7/2018

Problem Setting for Two-Class Classification

- m observations, each observation has n explanatory variables
- Each observation has a response value, $y_i \in \{-1, +1\}$ for $i = 1, \dots, m$

$$\begin{cases} i \in \Omega^+ & \text{if } y_i = +1 \\ i \in \Omega^- & \text{if } y_i = -1 \end{cases} \qquad \begin{cases} m_+ = |\Omega^+| \\ m_- = |\Omega^-| \end{cases}$$

- Explanatory matrix: $X \in \mathbb{R}^{m \times n}$
  - $X_i \in \mathbb{R}^n$ for $i = 1, \dots, m$ (row vector of X)
  - $x_j \in \mathbb{R}^m$ for $i = 1, \dots, n$ (column vector of X)
  - $x_{ij} \in \mathbb{R}$: $(i, j)^{th}$ element of X

	x_1	$\cdots$	x_j	$\cdots$	x_n	y
X_1						y_i
$\vdots$						$\vdots$
X_i			x_{ij}			y_i
$\vdots$						$\vdots$
X_m						y_i

Goal: Construct a classifier that accurately classifies positive or negative on unseen dataset

Classifier for two-class classification problems

Rule Function

Given $a, b \in \mathbb{R}^n$ with $a \le b$,

$$r_{(a,b)}(X_i) = \begin{cases} 1 & \text{if } a \le X_i \le b \text{ (componentwise)} \\ 0 & \text{otherwise} \end{cases}$$

Classifier constructed by a weighted combination of rules

$$\begin{aligned} f(X_i) &= \gamma_0 + \sum_{k \in K} \gamma_k r_k(X_i) \qquad\qquad \gamma_0, (\gamma_k)_{k \in K} \in \mathbb{R} \\ &= \gamma_0 + \sum_{k \in K} \left(\gamma_k^+ r_k^+(X_i) + \gamma_k^- r_k^-(X_i)\right) \end{aligned}$$

$\gamma_k > 0$: a observation covered by r_k votes to classify positive, and *vice versa*

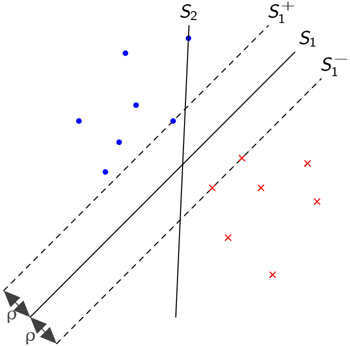
$$\text{Split } r_k(X_i) \text{ into } \begin{cases} r_k^+(X_i) = r_k(X_i) & \text{with } \gamma_k^+ \in \mathbb{R}_+ \\ r_k^-(X_i) = -r_k(X_i) & \text{with } \gamma_k^- \in \mathbb{R}_+ \end{cases}$$

Classification Rule

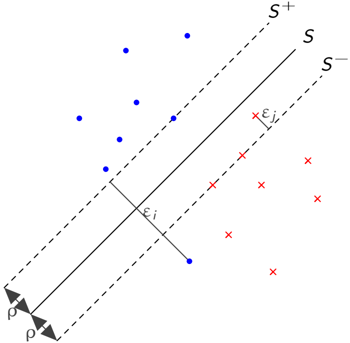
$$\text{Classify } \begin{cases} \text{positive} & \text{if } f(X_i) > 0 \text{ or } (f(X_i) = 0 \text{ and } m_+ \ge m_-) \\ \text{negative} & \text{if } f(X_i) < 0 \text{ or } (f(X_i) = 0 \text{ and } m_+ < m_-) \end{cases}$$

Margin Classifiers

Hard Margin Classifiers



Soft Margin Classifiers



Which one is a better classifier?
ρ: margin
ϵ_i: soft margin violation

Margin Classification Rule

$$\text{Classify} \begin{cases} \text{positive} & \text{if } f(X_i) \geq \rho \\ \text{negative} & \text{if } f(X_i) \leq \rho \end{cases}$$

LPBoost Master Problem

$$\min_{\rho,\epsilon,\gamma_0,\gamma^+,\gamma^-} \quad -\rho \quad + D \qquad \sum_{i=1}^{m} \epsilon_i^p$$

$$\text{s.t.} \qquad (\forall\, i) \quad y_i \left\{ \gamma_0 + \sum_{k \in K} \left(\gamma_k^+ r_k^+(X_i) + \gamma_k^- r_k^-(X_i) \right) \right\} + \epsilon_i \geq \rho$$

$$\sum_{k \in K} (\gamma_k^+ + \gamma_k^-) = 1$$

$$\epsilon \geq 0, \quad \gamma^+, \gamma^- \geq 0$$

- $D \in \mathbb{R}_+$: a constant parameter
  tradeoff between max margin and min soft margin violation
- $p = 1$ or 2

LPBoost Master Problem

$$\min_{\rho,\epsilon,\gamma_0,\gamma^+,\gamma^-} \quad -\rho \quad + D \quad \sum_{i=1}^{m} \epsilon_i^p$$

$$\text{s.t.} \qquad (\forall\, i) \quad y_i \underbrace{\left\{ \gamma_0 + \sum_{k \in K} \left(\gamma_k^+ r_k^+(X_i) + \gamma_k^- r_k^-(X_i) \right) \right\}}_{f(X_i)} + \epsilon_i \geq \rho$$

$$\sum_{k \in K} (\gamma_k^+ + \gamma_k^-) = 1$$

$$\epsilon \geq 0, \quad \gamma^+, \gamma^- \geq 0$$

- $D \in \mathbb{R}_+$: a constant parameter
  tradeoff between max margin and min soft margin violation
- $p = 1$ or 2

LPBoost Master Problem

$$\min_{\rho,\epsilon,\gamma_0,\gamma^+,\gamma^-} \underbrace{-\rho}_{\text{max margin}} + D \underbrace{\sum_{i=1}^{m} \epsilon_i^p}_{\text{min soft margin violation}}$$

$$\text{s.t.} \quad (\forall\, i) \quad y_i \underbrace{\left\{\gamma_0 + \sum_{k \in K} \left(\gamma_k^+ r_k^+(X_i) + \gamma_k^- r_k^-(X_i)\right)\right\}}_{f(X_i)} + \epsilon_i \geq \rho$$

$$\sum_{k \in K} (\gamma_k^+ + \gamma_k^-) = 1$$

$$\epsilon \geq 0, \quad \gamma^+, \gamma^- \geq 0$$

- $D \in \mathbb{R}_+$: a constant parameter
  tradeoff between max margin and min soft margin violation
- $p = 1$ or 2

LPBoost Master Problem

$$\min_{\rho,\epsilon,\gamma_0,\gamma^+,\gamma^-} \underbrace{-\rho}_{\text{max margin}} + D \underbrace{\sum_{i=1}^{m} \epsilon_i^p}_{\text{min soft margin violation}}$$

Dual Variables

s.t.

$$(\forall i) \quad \underbrace{y_i \left\{ \gamma_0 + \sum_{k \in K} \left(\gamma_k^+ r_k^+(X_i) + \gamma_k^- r_k^-(X_i) \right) \right\}}_{f(X_i)} + \epsilon_i \geq \rho \quad \leftrightarrow \mu_i \in \mathbb{R}_+$$

$$\sum_{k \in K} (\gamma_k^+ + \gamma_k^-) = 1 \quad \leftrightarrow \alpha \in \mathbb{R}$$

$$\epsilon \geq 0, \quad \gamma^+, \gamma^- \geq 0$$

- $D \in \mathbb{R}_+$: a constant parameter
  tradeoff between max margin and min soft margin violation
- $p = 1$ or 2

LPBoost Master Problem

$$\min_{\rho,\epsilon,\gamma_0,\gamma^+,\gamma^-} \quad \underbrace{-\rho}_{\text{max margin}} \quad + D \quad \underbrace{\sum_{i=1}^{m} \epsilon_i^p}_{\text{min soft margin violation}} \qquad \text{Dual Variables}$$

$$\text{s.t.} \qquad (\forall\, i) \quad y_i \underbrace{\left\{\gamma_0 + \sum_{k\in K} \left(\gamma_k^+ r_k^+(X_i) + \gamma_k^- r_k^-(X_i)\right)\right\}}_{f(X_i)} + \epsilon_i \geq \rho \leftrightarrow \mu_i \in \mathbb{R}_+$$

$$\sum_{k\in K} (\gamma_k^+ + \gamma_k^-) = 1 \qquad \leftrightarrow \alpha \in \mathbb{R}$$

$$\epsilon \geq 0, \quad \gamma^+, \gamma^- \geq 0$$

- $D \in \mathbb{R}_+$: a constant parameter
  tradeoff between max margin and min soft margin violation
- $p = 1$ or 2
- $m + 1$ constraints (except nonnegativity)
- $2 + m + 2|K|$ variables ($|K| \geq 3^n$)
- Large number of variables $\Rightarrow$ solve by **column generation**

LPBoost Pricing Problem

Link: Dual LPBR

- Finding the smallest reduced cost (most violated dual constraints)

LPBoost Pricing Problem

$$
\begin{aligned}
z^* &= \max_{k \in K} \left\{ \sum_{i=1}^{m} r_k^+(X_i) y_i \mu_i, \sum_{i=1}^{m} r_k^-(X_i) y_i \mu_i \right\} \\
&= \max_{k \in K} \left\{ \sum_{i=1}^{m} r_k^+(X_i) w_i, \sum_{i=1}^{m} r_k^-(X_i) w_i \right\} \\
&= \max_{k \in K} \left\{ \sum_{i}^{m} r_k(X_i) w_i, -\sum_{i}^{m} r_k(X_i) w_i \right\} \\
&= \max_{k \in K} \left| \sum_{i}^{m} r_k(X_i) w_i \right|
\end{aligned}
$$

- $w_i = y_i \mu_i$
  $$= \begin{cases} \mu_i & \text{if } i \in \Omega^+ \\ -\mu_i & \text{if } i \in \Omega^- \end{cases}$$
  $i = 1, \ldots, m$
- Stopping Condition: $z^* \leq -\alpha$

- This is RMA problem and solved by our solver
- LPBR (LPBoost with RMA)
- Maximizing the first term is finding the positive box r_k^+ s.t. the sum of the misclassification weights for covered positive observations most greatly exceeds the sum of the misclassification weights for covered negative observations

LPBR: Implementation of Column Generation

Restricted Master Problem: Initially, empty K

Pricing Problem: Initially, $\mu_i = \dfrac{1}{m}$ (Optional: Insert initial 1D rules for all attributes)

Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$z^* = \max_{k \in K} \left\{ \sum_{i=1}^{m} r_k^+(X_i) y_i \mu_i, \sum_{i=1}^{m} r_k^-(X_i) y_i \mu_i \right\}$$

LPBR: Implementation of Column Generation

Restricted Master Problem: Initially, empty K

A set of rules $r_k(\cdot)$
(can add multiple rules) $\Uparrow$

Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < -\alpha$ for Parallel Branch-and-Bound

LPBR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$\rho \in \mathbb{R}, \epsilon_i, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \quad r_k = \{r_k^+ \text{ or } -r_k^-\}$

$\gamma_k = \gamma_k^+ - \gamma_k^- \quad \hat{\gamma}_k = \gamma_k^+ + \gamma_k^-$

$$
\begin{array}{llllll}
\min & D\epsilon_1 \cdots +D\epsilon_m & -\rho & & & \\
\text{s.t.} & \epsilon_1 & -\rho & +r_1(X_1)\gamma_1 & & \geq 0 \\
 & \ddots & \vdots & \vdots & & \vdots \\
 & \quad +\epsilon_m & -\rho & +r_1(X_m)\gamma_1 & & \geq 0 \\
 & & & \hat{\gamma}_1 & & = 1
\end{array}
$$

A set of rules $r_k(\cdot)$
(can add multiple rules) $\Uparrow$

Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < -\alpha$ for Parallel Branch-and-Bound

LPBR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$\rho \in \mathbb{R}, \epsilon_i, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad r_k = \{r_k^+ \text{ or } -r_k^-\}$

$\gamma_k = \gamma_k^+ - \gamma_k^- \qquad \hat{\gamma}_k = \gamma_k^+ + \gamma_k^-$

$$\begin{array}{llllll}
\min & D\epsilon_1 \cdots +D\epsilon_m & -\rho & & & \\
\text{s.t.} & \epsilon_1 & -\rho & +r_1(X_1)\gamma_1 & & \geq 0 \\
& \ddots & \vdots & \vdots & & \vdots \\
& +\epsilon_m & -\rho & +r_1(X_m)\gamma_1 & & \geq 0 \\
& & & \hat{\gamma}_1 & & = 1
\end{array}$$

A set of rules $r_k(\cdot)$
(can add multiple rules) $\Uparrow$

$\Downarrow$ Optimal Dual Variables
$\mu \in \mathbb{R}_+^m$, $\alpha \in \mathbb{R}$

Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < -\alpha$ for Parallel Branch-and-Bound

LPBR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$\rho \in \mathbb{R}, \epsilon_i, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \quad r_k = \{r_k^+ \text{ or } -r_k^-\}$

$\gamma_k = \gamma_k^+ - \gamma_k^- \quad \hat{\gamma}_k = \gamma_k^+ + \gamma_k^-$

$$
\begin{array}{llllll}
\min & D\epsilon_1 \cdots +D\epsilon_m & -\rho & & & \\
\text{s.t.} & \epsilon_1 & -\rho & +r_1(X_1)\gamma_1 & & \geq 0 \\
& \ddots & \vdots & \vdots & & \vdots \\
& +\epsilon_m & -\rho & +r_1(X_m)\gamma_1 & & \geq 0 \\
& & & \hat{\gamma}_1 & & = 1
\end{array}
$$

A set of rules $r_k(\cdot)$ (can add multiple rules) $\Uparrow$

$\Downarrow$ Optimal Dual Variables $\mu \in \mathbb{R}_+^m$, $\alpha \in \mathbb{R}$

Pricing Problem:

Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$z^* = \max_{k \in K} \left\{ \sum_{i=1}^{m} r_k^+(X_i) y_i \mu_i, \sum_{i=1}^{m} r_k^-(X_i) y_i \mu_i \right\}$$

Stop if $z^* < -\alpha$ for Parallel Branch-and-Bound

LPBR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$\rho \in \mathbb{R}, \epsilon_i, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad r_k = \{r_k^+ \text{ or } -r_k^-\}$

$\gamma_k = \gamma_k^+ - \gamma_k^- \qquad \hat{\gamma}_k = \gamma_k^+ + \gamma_k^-$

$$\begin{array}{llllll}
\min & D\epsilon_1 \cdots + D\epsilon_m & -\rho & & & \\
\text{s.t.} & \epsilon_1 & -\rho & +r_1(X_1)\gamma_1 & & \geq 0 \\
 & \ddots & \vdots & \vdots & & \vdots \\
 & +\epsilon_m & -\rho & +r_1(X_m)\gamma_1 & & \geq 0 \\
 & & & \hat{\gamma}_1 & & = 1
\end{array}$$

A set of rules $r_k(\cdot)$ (can add multiple rules) $\Uparrow$

$\Downarrow$ Optimal Dual Variables $\mu \in \mathbb{R}_+^m$, $\alpha \in \mathbb{R}$

Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < -\alpha$ for Parallel Branch-and-Bound

LPBR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$\rho \in \mathbb{R}, \epsilon_i, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad r_k = \{r_k^+ \text{ or } -r_k^-\}$

$\gamma_k = \gamma_k^+ - \gamma_k^- \qquad \hat{\gamma}_k = \gamma_k^+ + \gamma_k^-$

$$
\begin{array}{llllllll}
\min & D\epsilon_1 & \cdots & +D\epsilon_m & -\rho & & & \\
\text{s.t.} & \epsilon_1 & & & -\rho & +r_1(X_1)\gamma_1 & +r_2(X_1)\gamma_2 & \geq 0 \\
 & & \ddots & & \vdots & \vdots & \vdots & \vdots \\
 & & & +\epsilon_m & -\rho & +r_1(X_m)\gamma_1 & +r_2(X_m)\gamma_2 & \geq 0 \\
 & & & & & \hat{\gamma}_1 & +\hat{\gamma}_2 & = 1
\end{array}
$$

A set of rules $r_k(\cdot)$ (can add multiple rules) $\Uparrow$

$\Downarrow$ Optimal Dual Variables $\mu \in \mathbb{R}_+^m$, $\alpha \in \mathbb{R}$

Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < -\alpha$ for Parallel Branch-and-Bound

LPBR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$$\rho \in \mathbb{R}, \epsilon_i, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad r_k = \{r_k^+ \text{ or } -r_k^-\}$$

$$\gamma_k = \gamma_k^+ - \gamma_k^- \qquad \hat{\gamma}_k = \gamma_k^+ + \gamma_k^-$$

$$\begin{array}{llllllll}
\min & D\epsilon_1 \cdots +D\epsilon_m & -\rho & & & & \\
\text{s.t.} & \epsilon_1 & -\rho & +r_1(X_1)\gamma_1 & +r_2(X_1)\gamma_2 & \cdots & \geq 0 \\
& \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
& +\epsilon_m & -\rho & +r_1(X_m)\gamma_1 & +r_2(X_m)\gamma_2 & \cdots & \geq 0 \\
& & & \hat{\gamma}_1 & +\hat{\gamma}_2 & \cdots & = 1
\end{array}$$

A set of rules $r_k(\cdot)$ (can add multiple rules) $\Uparrow$

$\Downarrow$ Optimal Dual Variables $\mu \in \mathbb{R}_+^m$, $\alpha \in \mathbb{R}$

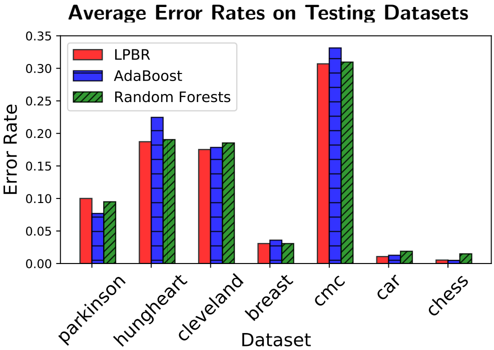
Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < -\alpha$ for Parallel Branch-and-Bound

Result: LPBR has lower error rates on testing datasets

- Bi-level cross-validation to choose $D \in \{0.0001, 0.001, 0.005, 0.01\}$ with $p = 2$
- Twice of 5-fold outer and 3-fold inner cross-validations
- 100 column generation iterations (add one rule per iteration)
- Solve the subproblems by greedy RMA in general
- If the objective value of the greedy RMA satisfies the stopping condition, then solve the subproblems by the parallel branch-and-bound procedure

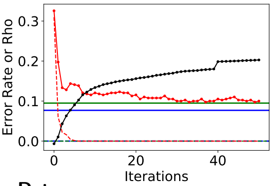


- Random Forests (R package: randomForest) with 100 trees
- AdaBoost (R package: fastAdaboost) with 100 iterations

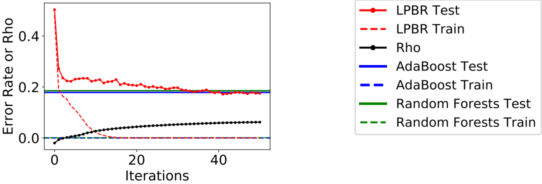
Result: Error rates decreases in each column generation iteration

- Initially generated one-dimensional greedy box-based rules
- First few rules has larger effect to minimize classification error rates
- Testing error rates decrease by increasing ρ (margin) in each iteration

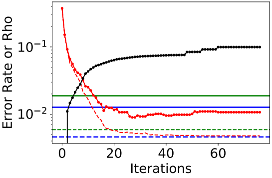
Data: parkinson



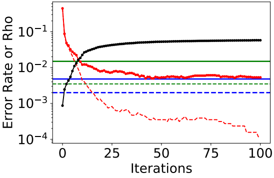
Data: cleveland



Data: car



Data: chess



Second RMA Application: Regression Model

- m observations, each observation has n explanatory variables
- Each observation has a response value, $y_i \in \mathbb{R}$ for $i = 1, \ldots, m$
- Goal: Construct a prediction function $f(\cdot)$ for unseen data $(X_{i'}, y_{i'})$ such that minimizing prediction error ($\min\{f(X_{i'}) - y_{i'}\}$)
- Do not have *a priori* functional form for f

Penalized Regression with Rules

Rule Function

Given $a, b \in \mathbb{R}^n$ with $a \leq b$, $r_{(a,b)}(X_i) = \begin{cases} 1 & \text{if } a \leq X_i \leq b \text{ (componentwise)} \\ 0 & \text{otherwise} \end{cases}$

Prediction Function (Linear & Rule Regression)

For some $\beta_0, \beta_1, \ldots, \beta_n, (\gamma_k)_{k \in K} \in \mathbb{R}$,

$$f(X_i) = f_{\beta_0,\beta,\gamma}(X_i) = \beta_0 + \sum_{j=1}^{n} \beta_j x_{ij} + \sum_{k \in K} \gamma_k r_k(X_i)$$

Rule-Enhanced Penalized Regression (REPR)

For $\beta = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}^{|K|}$,

$$\min_{\beta_0,\beta,\gamma} \left\{ \underbrace{\sum_{i=1}^{m} \left| f_{\beta_0,\beta,\gamma}(X_i) - y_i \right|^p}_{\text{min prediction error}} \quad \underbrace{+C \, \|\beta\|_1}_{\text{penalize linear coefficients}} \quad \underbrace{+E \, \|\gamma\|_1}_{\text{penalize rule coefficients}} \right\}$$

$p \in \{1, 2\}$ and $C, E \geq 0$ are scalar parameters

- For $p = 2$ and $C = E > 0$, this model is the classic LASSO
- Split

$$\begin{cases} \beta = \beta^+ - \beta^- & \text{with } \beta^+, \beta^- \geq 0 \\ \gamma = \gamma^+ - \gamma^- & \text{with } \gamma^+, \gamma^- \geq 0 \end{cases}$$

REPR Master Problem

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{m} \epsilon_i^p \quad + C \sum_{j=1}^{n} (\beta_j^+ + \beta_j^-) + E \sum_{k \in K} (\gamma_k^+ + \gamma_k^-) \\
\text{s.t.} \quad & \beta_0 + X_i^\top (\beta^+ - \beta^-) + \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-) - \epsilon_i \leq \quad y_i \\
& (\forall\, i) \;\; -\beta_0 - X_i^\top (\beta^+ - \beta^-) - \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-) - \epsilon_i \leq -y_i \\
& \epsilon \geq 0, \qquad \beta^+, \beta^- \geq 0, \qquad \gamma^+, \gamma^- \geq 0
\end{aligned}
$$

REPR Master Problem

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{m} \epsilon_i^p + C \sum_{j=1}^{n} (\beta_j^+ + \beta_j^-) + E \sum_{k \in K} (\gamma_k^+ + \gamma_k^-) \\
\text{s.t.} \quad & \underbrace{\beta_0 + X_i^\top (\beta^+ - \beta^-) + \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-)}_{f_{\beta_0,\beta,\gamma}(X_i)} - \epsilon_i \le \quad y_i \\
& (\forall\, i) \; \underbrace{-\beta_0 - X_i^\top (\beta^+ - \beta^-) - \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-)}_{-f_{\beta_0,\beta,\gamma}(X_i)} - \epsilon_i \le -y_i \\
& \epsilon \ge 0, \qquad \beta^+, \beta^- \ge 0, \qquad \gamma^+, \gamma^- \ge 0
\end{aligned}
$$

Overestimation constraints: $f_{\beta_0,\beta,\gamma}(X_i) - y_i \le \epsilon_i$

Underestimation constraints: $f_{\beta_0,\beta,\gamma}(X_i) - y_i \ge -\epsilon_i$

REPR Master Problem

$$\min \quad \underbrace{\sum_{i=1}^{m} \epsilon_i^p}_{\text{min prediction error}} \quad \underbrace{+ C \sum_{j=1}^{n} (\beta_j^+ + \beta_j^-)}_{\text{penalize linear coefficients}} \underbrace{+ E \sum_{k \in K} (\gamma_k^+ + \gamma_k^-)}_{\text{penalize rule coefficients}}$$

$$\text{s. t.} \quad \underbrace{\beta_0 + X_i^\top (\beta^+ - \beta^-) + \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-)}_{f_{\beta_0, \beta, \gamma}(X_i)} - \epsilon_i \leq \quad y_i$$

$$(\forall\, i) \ \underbrace{-\beta_0 - X_i^\top (\beta^+ - \beta^-) - \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-)}_{-f_{\beta_0, \beta, \gamma}(X_i)} - \epsilon_i \leq -y_i$$

$$\epsilon \geq 0, \qquad \beta^+, \beta^- \geq 0, \qquad \gamma^+, \gamma^- \geq 0$$

Overestimation constraints: $f_{\beta_0,\beta,\gamma}(X_i) - y_i \leq \quad \epsilon_i$

Underestimation constraints: $f_{\beta_0,\beta,\gamma}(X_i) - y_i \geq -\epsilon_i$

REPR Master Problem

$$\min \quad \underbrace{\sum_{i=1}^{m} \epsilon_i^p}_{\text{min prediction error}} \quad \underbrace{+ C \sum_{j=1}^{n} (\beta_j^+ + \beta_j^-)}_{\text{penalize linear coefficients}} \underbrace{+ E \sum_{k \in K} (\gamma_k^+ + \gamma_k^-)}_{\text{penalize rule coefficients}}$$

Dual Variables

$$\text{s.t.} \quad \underbrace{\beta_0 + X_i^\top (\beta^+ - \beta^-) + \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-)}_{f_{\beta_0, \beta, \gamma}(X_i)} - \epsilon_i \leq \quad y_i \qquad \leftrightarrow \mu_i \in \mathbb{R}_+$$

$$(\forall\, i) \quad \underbrace{-\beta_0 - X_i^\top (\beta^+ - \beta^-) - \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-)}_{-f_{\beta_0, \beta, \gamma}(X_i)} - \epsilon_i \leq -y_i \qquad \leftrightarrow \nu_i \in \mathbb{R}_+$$

$$\epsilon \geq 0, \qquad \beta^+, \beta^- \geq 0, \qquad \gamma^+, \gamma^- \geq 0$$

Overestimation constraints: $f_{\beta_0,\beta,\gamma}(X_i) - y_i \leq \epsilon_i$

Underestimation constraints: $f_{\beta_0,\beta,\gamma}(X_i) - y_i \geq -\epsilon_i$

Overestimated: $\mu_i > 0, \nu_i = 0$

Underestimated: $\mu_i = 0, \nu_i > 0$

REPR Master Problem

$$\min \quad \underbrace{\sum_{i=1}^{m} \epsilon_i^p}_{\text{min prediction error}} \; \underbrace{+\, C \sum_{j=1}^{n} (\beta_j^+ + \beta_j^-)}_{\text{penalize linear coefficients}} \; \underbrace{+\, E \sum_{k \in K} (\gamma_k^+ + \gamma_k^-)}_{\text{penalize rule coefficients}} \qquad \text{Dual Variables}$$

$$\text{s.t.} \quad \underbrace{\beta_0 + X_i^\top(\beta^+ - \beta^-) + \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-)}_{f_{\beta_0,\beta,\gamma}(X_i)} - \epsilon_i \le \quad y_i \qquad \leftrightarrow \mu_i \in \mathbb{R}_+$$

$$(\forall\, i) \;\; \underbrace{-\beta_0 - X_i^\top(\beta^+ - \beta^-) - \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-)}_{-f_{\beta_0,\beta,\gamma}(X_i)} - \epsilon_i \le -y_i \qquad \leftrightarrow \nu_i \in \mathbb{R}_+$$

$$\epsilon \ge 0, \qquad \beta^+, \beta^- \ge 0, \qquad \gamma^+, \gamma^- \ge 0$$

Overestimation constraints: $f_{\beta_0,\beta,\gamma}(X_i) - y_i \le \quad \epsilon_i$

Underestimation constraints: $f_{\beta_0,\beta,\gamma}(X_i) - y_i \ge -\epsilon_i$

Overestimated: $\mu_i > 0, \nu_i = 0$

Underestimated: $\mu_i = 0, \nu_i > 0$

- $2m$ constraints (except nonnegativity)
- $1 + m + 2n + 2|K|$ variables ($|K| \ge 3^n$)
- Large number of variables $\Rightarrow$ solve by **column generation**

Pricing Problem and RMA Formulation

- Finding the smallest reduced cost (most violated dual constraints)

REPR Pricing Problem Link: Reduced Cost

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

- This is $\mathcal{NP}$-hard Rectangular Maximum Agreement (RMA) problem

Rectangular Maximum Agreement (RMA) Problem

$$\begin{array}{ll} \max\limits_{a,b \in \mathbb{R}^n} & \left| w\big(\mathrm{Cover}_X(a,b)\big) \right| \\ \mathrm{s.\,t.} & a, b \in \mathbb{R}^n \end{array}$$

- REPR pricing problem reduces to RMA with

$$\begin{cases} w_i & = \mu_i - \nu_i \quad (\forall\, i) \\ \Omega^+ & = \{ i \in \{1, \ldots, m\} \mid \mu_i \geq \nu_i \} \quad \text{Overestimated} \quad (\mu_i > 0, \nu_i = 0) \\ \Omega^- & = \{ i \in \{1, \ldots, m\} \mid \mu_i < \nu_i \} \quad \text{Underestimated} \ (\mu_i = 0, \nu_i > 0) \end{cases}$$

- Solve RMA by a specialized parallel branch-and-bound procedure

Some Insight into Column Generation

Link: REPR Dual

$$\begin{aligned}
\min \quad & \sum_{i=1}^{m} \epsilon_i^p + C \sum_{j=1}^{n} (\beta_j^+ + \beta_j^-) + E \sum_{k \in K} (\gamma_k^+ + \gamma_k^-) \\
\text{s.t.} \quad & (\forall i) \quad \beta_0 + X_i^\top (\beta^+ - \beta^-) + \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-) - \epsilon_i \le y_i \quad \longleftrightarrow \quad \mu_i \\
& (\forall i) \quad -X_i^\top (\beta^+ - \beta^-) - \sum_{k \in K} r_k(X_i)(\gamma_k^+ - \gamma_k^-) - \epsilon_i \le -y_i \quad \longleftrightarrow \quad \nu_i \\
& \epsilon \ge 0, \qquad \beta^+, \beta^- \ge 0, \qquad \gamma^+, \gamma^- \ge 0
\end{aligned}$$

$p = 1$ $\begin{cases} w_i = \mu_i = 1 & \text{when observation } i \text{ is } \textbf{overestimated} \\ w_i = -\nu_i = -1 & \text{when observation } i \text{ is } \textbf{underestimated} \end{cases}$

Pricing problem: find a box in which overestimated observations most outnumber underestimated ones or *vice versa*

$p = 2$ $\begin{cases} w_i = \mu_i = 2\epsilon_i & \text{when observation } i \text{ is } \textbf{overestimated} \\ w_i = -\nu_i = -2\epsilon_i & \text{when observation } i \text{ is } \textbf{underestimated} \end{cases}$

Pricing problem: find a box that the sum of overestimation errors most greatly exceeds the sum of underestimation errors or *vice versa*

REPR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI; Initially, empty K

$$\epsilon_i, \beta_j^+, \beta_j^-, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad \begin{aligned} \hat{\beta}_j &= \beta_j^+ + \beta_j^- \\ \beta_j &= \beta_j^+ - \beta_j^- \end{aligned}$$

$$\begin{array}{llllr}
\min & \epsilon_1^p \cdots + \epsilon_m^p & + C\hat{\beta}_1 \cdots + C\hat{\beta}_n & & \\
\text{s.t.} & \epsilon_1 & + \beta_0 + x_{11}\beta_1 \cdots + x_{1n}\beta_n & & \le y_1 \\
 & \ddots & \vdots & & \vdots \quad \vdots \\
 & + \epsilon_m & + \beta_0 + x_{m1}\beta_1 \cdots + x_{mn}\beta_n & & \le y_m \\
 & -\epsilon_1 & - \beta_0 - x_{11}\beta_1 \cdots - x_{1n}\beta_n & & \le -y_1 \\
 & \ddots & \vdots & & \vdots \quad \vdots \\
 & -\epsilon_m & - \beta_0 - x_{m1}\beta_1 \cdots - x_{mn}\beta_n & & \le -y_m
\end{array}$$

REPR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$$\epsilon_i, \beta_j^+, \beta_j^-, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad \begin{aligned}\hat{\beta}_j &= \beta_j^+ + \beta_j^- \\ \beta_j &= \beta_j^+ - \beta_j^-\end{aligned}$$

$$\begin{array}{llllll}
\min & \epsilon_1^p \cdots + \epsilon_m^p & + C\hat{\beta}_1 \cdots + C\hat{\beta}_n & & & \\
\text{s.t.} & \epsilon_1 & +\beta_0 + x_{11}\beta_1 \cdots + x_{1n}\beta_n & & & \le y_1 \\
& \ddots & \vdots & & & \vdots \quad \vdots \\
& \quad +\epsilon_m & +\beta_0 + x_{m1}\beta_1 \cdots + x_{mn}\beta_n & & & \le y_m \\
& -\epsilon_1 & -\beta_0 - x_{11}\beta_1 \cdots - x_{1n}\beta_n & & & \le -y_1 \\
& \ddots & \vdots & & & \vdots \quad \vdots \\
& \quad -\epsilon_m & -\beta_0 - x_{m1}\beta_1 \cdots - x_{mn}\beta_n & & & \le -y_m
\end{array}$$

$\Downarrow$ Optimal Dual Variables $\mu, \nu \in \mathbb{R}^m$

REPR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$$\epsilon_i, \beta_j^+, \beta_j^-, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad \hat{\beta}_j = \beta_j^+ + \beta_j^-, \quad \beta_j = \beta_j^+ - \beta_j^-$$

$$\begin{array}{llllll}
\min & \epsilon_1^p \cdots + \epsilon_m^p & & + C\hat{\beta}_1 \cdots + C\hat{\beta}_n & & \\
\text{s.t.} & \epsilon_1 & & + \beta_0 + x_{11}\beta_1 \cdots + x_{1n}\beta_n & & \leq y_1 \\
& \ddots & & \vdots & & \vdots \quad \vdots \\
& + \epsilon_m & & + \beta_0 + x_{m1}\beta_1 \cdots + x_{mn}\beta_n & & \leq y_m \\
& -\epsilon_1 & & -\beta_0 - x_{11}\beta_1 \cdots - x_{1n}\beta_n & & \leq -y_1 \\
& \ddots & & \vdots & & \vdots \quad \vdots \\
& - \epsilon_m & & -\beta_0 - x_{m1}\beta_1 \cdots - x_{mn}\beta_n & & \leq -y_m
\end{array}$$

$\Downarrow$ Optimal Dual Variables $\mu, \nu \in \mathbb{R}^m$

Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$z^* = \max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < E$ for Parallel Branch-and-Bound

REPR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$$\epsilon_i, \beta_j^+, \beta_j^-, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad \hat{\beta}_j = \beta_j^+ + \beta_j^- \quad \beta_j = \beta_j^+ - \beta_j^-$$

$$\begin{array}{llllll}
\min & \epsilon_1^p \cdots + \epsilon_m^p & + C\hat{\beta}_1 \cdots + C\hat{\beta}_n & & & \\
\text{s.t.} & \epsilon_1 & + \beta_0 + x_{11}\beta_1 \cdots + x_{1n}\beta_n & & & \leq y_1 \\
& \ddots & \vdots & & & \vdots \quad \vdots \\
& \quad + \epsilon_m & + \beta_0 + x_{m1}\beta_1 \cdots + x_{mn}\beta_n & & & \leq y_m \\
& -\epsilon_1 & - \beta_0 - x_{11}\beta_1 \cdots - x_{1n}\beta_n & & & \leq -y_1 \\
& \ddots & \vdots & & & \vdots \quad \vdots \\
& \quad -\epsilon_m & - \beta_0 - x_{m1}\beta_1 \cdots - x_{mn}\beta_n & & & \leq -y_m
\end{array}$$

A set of rules $r_k(\cdot)$ (can add multiple rules) $\Uparrow$

$\Downarrow$ Optimal Dual Variables $\mu, \nu \in \mathbb{R}^m$

Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < E$ for Parallel Branch-and-Bound

REPR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$$\epsilon_i, \beta_j^+, \beta_j^-, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad \begin{array}{l} \hat{\beta}_j = \beta_j^+ + \beta_j^- \\ \beta_j = \beta_j^+ - \beta_j^- \end{array} \qquad \begin{array}{l} \hat{\gamma}_j = \gamma_k^+ + \gamma_k^- \\ \gamma_k = \gamma_k^+ - \gamma_k^- \end{array}$$

$$\begin{array}{lllll}
\min & \epsilon_1^p \cdots + \epsilon_m^p & +C\hat{\beta}_1 \cdots +C\hat{\beta}_n & +D\hat{\gamma}_1 & \\
\text{s.t.} & \epsilon_1 & +\beta_0 + x_{11}\beta_1 \cdots + x_{1n}\beta_n & +r_1(X_1)\gamma_1 & \leq y_1 \\
 & \ddots & \vdots & \vdots & \vdots \quad \vdots \\
 & \qquad +\epsilon_m & +\beta_0 + x_{m1}\beta_1 \cdots + x_{mn}\beta_n & +r_1(X_m)\gamma_1 & \leq y_m \\
 & -\epsilon_1 & -\beta_0 - x_{11}\beta_1 \cdots - x_{1n}\beta_n & -r_1(X_1)\gamma_0 & \leq -y_1 \\
 & \ddots & \vdots & \vdots & \vdots \quad \vdots \\
 & \qquad -\epsilon_m & -\beta_0 - x_{m1}\beta_1 \cdots - x_{mn}\beta_n & -r_1(X_m)\gamma_1 & \leq -y_m
\end{array}$$

A set of rules $r_k(\cdot)$ (can add multiple rules) $\Uparrow$

$\Downarrow$ Optimal Dual Variables $\mu, \nu \in \mathbb{R}^m$

Pricing Problem:

Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < E$ for Parallel Branch-and-Bound

REPR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$$\epsilon_i, \beta_j^+, \beta_j^-, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad \hat{\beta}_j = \beta_j^+ + \beta_j^- \qquad \hat{\gamma}_j = \gamma_k^+ + \gamma_k^-$$
$$\beta_j = \beta_j^+ - \beta_j^- \qquad \gamma_k = \gamma_k^+ - \gamma_k^-$$

$$\begin{array}{lllll}
\min & \epsilon_1^p \cdots + \epsilon_m^p & +C\hat{\beta}_1 \cdots +C\hat{\beta}_n & +D\hat{\gamma}_1 & \\
\text{s.t.} & \epsilon_1 & +\beta_0 + x_{11}\beta_1 \cdots + x_{1n}\beta_n & +r_1(X_1)\gamma_1 & \leq y_1 \\
& \quad \ddots & \vdots & \vdots & \vdots \quad \vdots \\
& \qquad +\epsilon_m & +\beta_0 + x_{m1}\beta_1 \cdots + x_{mn}\beta_n & +r_1(X_m)\gamma_1 & \leq y_m \\
& -\epsilon_1 & -\beta_0 - x_{11}\beta_1 \cdots - x_{1n}\beta_n & -r_1(X_1)\gamma_0 & \leq -y_1 \\
& \quad \ddots & \vdots & \vdots & \vdots \quad \vdots \\
& \qquad -\epsilon_m & -\beta_0 - x_{m1}\beta_1 \cdots - x_{mn}\beta_n & -r_1(X_m)\gamma_1 & \leq -y_m
\end{array}$$

A set of rules $r_k(\cdot)$ (can add multiple rules) $\Uparrow$

$\Downarrow$ Optimal Dual Variables $\mu, \nu \in \mathbb{R}^m$

Pricing Problem:

Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < E$ for Parallel Branch-and-Bound

REPR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$$\epsilon_i, \beta_j^+, \beta_j^-, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad \hat{\beta}_j = \beta_j^+ + \beta_j^- \quad \hat{\gamma}_j = \gamma_k^+ + \gamma_k^-$$
$$\beta_j = \beta_j^+ - \beta_j^- \quad \gamma_k = \gamma_k^+ - \gamma_k^-$$

$$\begin{array}{lllll}
\min & \epsilon_1^p \cdots + \epsilon_m^p & + C\hat{\beta}_1 \cdots + C\hat{\beta}_n & + D\hat{\gamma}_1 & \\
\text{s. t.} & \epsilon_1 & + \beta_0 + x_{11}\beta_1 \cdots + x_{1n}\beta_n & + r_1(X_1)\gamma_1 & \le y_1 \\
 & \ddots & \vdots & \vdots & \vdots \quad \vdots \\
 & + \epsilon_m & + \beta_0 + x_{m1}\beta_1 \cdots + x_{mn}\beta_n & + r_1(X_m)\gamma_1 & \le y_m \\
 & -\epsilon_1 & - \beta_0 - x_{11}\beta_1 \cdots - x_{1n}\beta_n & - r_1(X_1)\gamma_0 & \le -y_1 \\
 & \ddots & \vdots & \vdots & \vdots \quad \vdots \\
 & -\epsilon_m & - \beta_0 - x_{m1}\beta_1 \cdots - x_{mn}\beta_n & - r_1(X_m)\gamma_1 & \le -y_m
\end{array}$$

A set of rules $r_k(\cdot)$ (can add multiple rules) $\Uparrow$ $\Downarrow$ Optimal Dual Variables $\mu, \nu \in \mathbb{R}^m$

Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$z^* = \max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < E$ for Parallel Branch-and-Bound

REPR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$$\epsilon_i, \beta_j^+, \beta_j^-, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad \hat{\beta}_j = \beta_j^+ + \beta_j^- \quad \hat{\gamma}_j = \gamma_k^+ + \gamma_k^-$$
$$\beta_j = \beta_j^+ - \beta_j^- \quad \gamma_k = \gamma_k^+ - \gamma_k^-$$

$$\begin{array}{llllll}
\min & \epsilon_1^p \cdots +\epsilon_m^p & +C\hat{\beta}_1 \cdots +C\hat{\beta}_n & +D\hat{\gamma}_1 & & \\
\text{s.t.} & \epsilon_1 & +\beta_0 + x_{11}\beta_1 \cdots + x_{1n}\beta_n & +r_1(X_1)\gamma_1 & \leq & y_1 \\
& \ddots & \vdots & \vdots & \vdots & \vdots \\
& +\epsilon_m & +\beta_0 + x_{m1}\beta_1 \cdots + x_{mn}\beta_n & +r_1(X_m)\gamma_1 & \leq & y_m \\
& -\epsilon_1 & -\beta_0 - x_{11}\beta_1 \cdots - x_{1n}\beta_n & -r_1(X_1)\gamma_0 & \leq & -y_1 \\
& \ddots & \vdots & \vdots & \vdots & \vdots \\
& -\epsilon_m & -\beta_0 - x_{m1}\beta_1 \cdots - x_{mn}\beta_n & -r_1(X_m)\gamma_1 & \leq & -y_m
\end{array}$$

A set of rules $r_k(\cdot)$
(can add multiple rules) $\Uparrow$

$\Downarrow$ Optimal Dual Variables
$\mu, \nu \in \mathbb{R}^m$

Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < E$ for Parallel Branch-and-Bound

REPR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$$\epsilon_i, \beta_j^+, \beta_j^-, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad \hat{\beta}_j = \beta_j^+ + \beta_j^- \quad \hat{\gamma}_j = \gamma_k^+ + \gamma_k^-$$
$$\beta_j = \beta_j^+ - \beta_j^- \quad \gamma_k = \gamma_k^+ - \gamma_k^-$$

$$\begin{array}{llllll}
\min & \epsilon_1^p \cdots + \epsilon_m^p & +C\hat{\beta}_1 \cdots +C\hat{\beta}_n & +D\hat{\gamma}_1 & +D\hat{\gamma}_2 & \\
\text{s.t.} & \epsilon_1 & +\beta_0 + x_{11}\beta_1 \cdots + x_{1n}\beta_n & +r_1(X_1)\gamma_1 & +r_2(X_1)\gamma_2 & \leq y_1 \\
& \ddots & \vdots & \vdots & \vdots & \vdots \quad \vdots \\
& +\epsilon_m & +\beta_0 + x_{m1}\beta_1 \cdots + x_{mn}\beta_n & +r_1(X_m)\gamma_1 & +r_2(X_m)\gamma_2 & \leq y_m \\
& -\epsilon_1 & -\beta_0 - x_{11}\beta_1 \cdots - x_{1n}\beta_n & -r_1(X_1)\gamma_0 & -r_2(X_1)\gamma_2 & \leq -y_1 \\
& \ddots & \vdots & \vdots & \vdots & \vdots \quad \vdots \\
& -\epsilon_m & -\beta_0 - x_{m1}\beta_1 \cdots - x_{mn}\beta_n & -r_1(X_m)\gamma_1 & -r_2(X_m)\gamma_2 & \leq -y_m
\end{array}$$

A set of rules $r_k(\cdot)$ (can add multiple rules) $\Uparrow$ $\Downarrow$ Optimal Dual Variables $\mu, \nu \in \mathbb{R}^m$

Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < E$ for Parallel Branch-and-Bound

REPR: Implementation of Column Generation

Restricted Master Problem: Solved by GUROBI

$$\epsilon_i, \beta_j^+, \beta_j^-, \gamma_k^+, \gamma_k^- \in \mathbb{R}^+ \qquad \hat{\beta}_j = \beta_j^+ + \beta_j^- \qquad \hat{\gamma}_j = \gamma_k^+ + \gamma_k^-$$
$$\beta_j = \beta_j^+ - \beta_j^- \qquad \gamma_k = \gamma_k^+ - \gamma_k^-$$

$$\begin{array}{lllllll}
\min & \epsilon_1^p \cdots +\epsilon_m^p & +C\hat{\beta}_1 \cdots +C\hat{\beta}_n & +D\hat{\gamma}_1 & +D\hat{\gamma}_2 & \cdots & \\
\text{s.t.} & \epsilon_1 & +\beta_0 + x_{11}\beta_1 \cdots + x_{1n}\beta_n & +r_1(X_1)\gamma_1 & +r_2(X_1)\gamma_2 & \cdots & \leq y_1 \\
 & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \quad \vdots \\
 & +\epsilon_m & +\beta_0 + x_{m1}\beta_1 \cdots + x_{mn}\beta_n & +r_1(X_m)\gamma_1 & +r_2(X_m)\gamma_2 & \cdots & \leq y_m \\
 & -\epsilon_1 & -\beta_0 - x_{11}\beta_1 \cdots - x_{1n}\beta_n & -r_1(X_1)\gamma_0 & -r_2(X_1)\gamma_2 & \cdots & \leq -y_1 \\
 & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \quad \vdots \\
 & -\epsilon_m & -\beta_0 - x_{m1}\beta_1 \cdots - x_{mn}\beta_n & -r_1(X_m)\gamma_1 & -r_2(X_m)\gamma_2 & \cdots & \leq -y_m
\end{array}$$

A set of rules $r_k(\cdot)$ (can add multiple rules) $\Uparrow$ $\Downarrow$ Optimal Dual Variables $\mu, \nu \in \mathbb{R}^m$

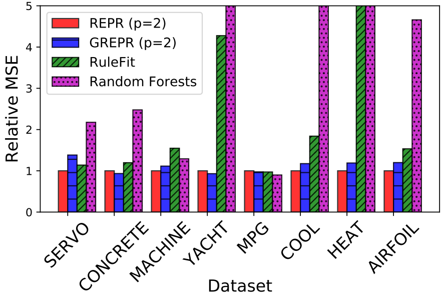
Pricing Problem:
Solved by Parallel Branch-and-Bound using PEBBL OR Greedy Heuristic

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Stop if $z^* < E$ for Parallel Branch-and-Bound

Result: REPR has lower relative MSE on testing datasets

- Bi-level cross-validation (inner cross-validation to choose $C = E$), $p = 2$
- 150 column generation iterations (add one rule per iteration)
- **REPR**: Pricing by parallel branch and bound
- **GREPR**: Pricing by greedy heuristic (serial)

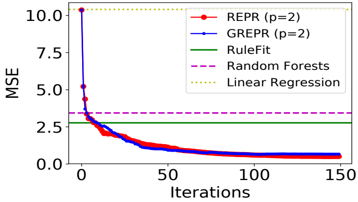


- REPR average MSE normalized to 1
- Some bars truncated
- Random Forests, RuleFit, and Linear Regression
  were implemented by their R packages with default setting (default parameters)

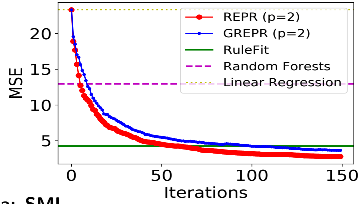
Result: MSE decreases in each column generation iteration (No obvious overfitting)

- First iteration of REPR is the same as LASSO
- First few rules have larger effect to minimize MSE on testing datasets

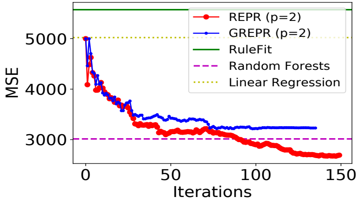
Data: HEAT



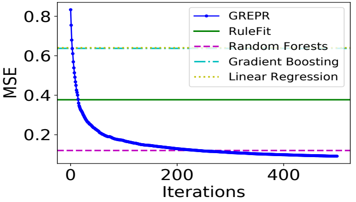
Data: AIRFOIL



Data: MACHINE



Data: SML



Conclusion and Planned Improvements/Extensions

Conclusion

- LPBR constructs competitive classification models
- REPR makes more accurate and robust predictions than the other methods such as RuleFit and Random Forests

Improvement

- Try using larger datasets
- Dimensional reduction
- More parallel procedures
- Smarter parameter choices

LPBR Dual Formulation

Link: LPBR Pricing

$p = 1$

$$
\begin{aligned}
\max_{\mu \geq 0,\, \alpha} \quad & \alpha \\
\text{s.t.} \quad & -\sum_{i=1}^{m} \mu_i y_i r_k^+(X_i) \geq \alpha, && k \in K \\
& -\sum_{i=1}^{m} \mu_i y_i r_k^-(X_i) \geq \alpha, && k \in K \\
& \sum_{i=1}^{m} y_i \mu_i = 0, \quad \sum_{i=1}^{m} \mu_i = 1 \\
& 0 \leq \mu_i \leq D, && i = 1, \ldots, m
\end{aligned}
$$

$p = 2$

$$
\begin{aligned}
\max_{\mu \geq 0,\, \alpha} \quad & \alpha + \sum_{i=1}^{m} \left(\frac{\mu_i}{2} - \frac{\mu_i^2}{2D} \right) \\
\text{s.t.} \quad & -\sum_{i=1}^{m} \mu_i y_i r_k^+(X_i) \geq \alpha, && k \in K \\
& -\sum_{i=1}^{m} \mu_i y_i r_k^-(X_i) \geq \alpha, && k \in K \\
& \sum_{i=1}^{m} y_i \mu_i = 0, \quad \sum_{i=1}^{m} \mu_i = 1
\end{aligned}
$$

REPR Pricing Problem

Link: Pricing Problem

- **Reduced cost**: rate at which optimal objective changes as you increase the value of a variable

Reduced Cost

$$\mathrm{rc}[\gamma_k^+] = E - \sum_{i=1}^{m} r_k(X_i)\nu_i + \sum_{i=1}^{m} r_k(X_i)\mu_i$$

$$\mathrm{rc}[\gamma_k^-] = E + \sum_{i=1}^{m} r_k(X_i)\nu_i - \sum_{i=1}^{m} r_k(X_i)\mu_i$$

- Finding the smallest reduced cost
  - most violated dual constraints
  - variable whose introduction improves the objective fastest

REPR Pricing Problem

$$\max_{k \in K} \left| \sum_{i=1}^{m} r_k(X_i)(\nu_i - \mu_i) \right|$$

Dual Formulation of RPER

Link: Insight

$p = 1$

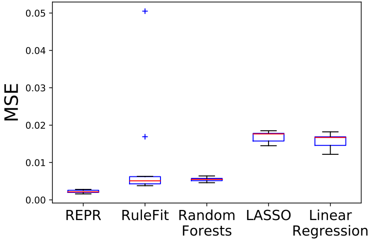
$$\begin{aligned} \max_{\mu,\nu \geq 0} \quad & \sum_{i=1}^{m} y_i(\nu_i - \mu_i) & \\ \text{s.t.} \quad & \sum_{i=1}^{m} (\mu_i - \nu_i) = 0 & \\ & \mu_i + \nu_i \leq 1, & i = 1, \ldots, m \\ & \left| \sum_{i=1}^{m} \nu_i x_{ij} - \sum_{i=1}^{m} \mu_i x_{ij} \right| \leq C, & j = 1, \ldots n \\ & \left| \sum_{i=1}^{m} \nu_i r_k(X_i) - \sum_{i=1}^{m} \mu_i r_k(X_i) \right| \leq E, & k \in K \end{aligned}$$

$p = 2$

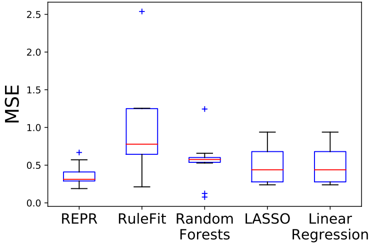
$$\begin{aligned} \max_{\mu,\nu \geq 0} \quad & \sum_{i=1}^{m} y_i(\nu_i - \mu_i) - \frac{1}{4} \sum_{i=1}^{m} (\nu_i + \mu_i)^2 & \\ \text{s.t.} \quad & \sum_{i=1}^{m} (\mu_i - \nu_i) = 0 & \\ & \left| \sum_{i=1}^{m} \nu_i x_{ij} - \sum_{i=1}^{m} \mu_i x_{ij} \right| \leq C, & j = 1, \ldots n \\ & \left| \sum_{i=1}^{m} \nu_i r_k(X_i) - \sum_{i=1}^{m} \mu_i r_k(X_i) \right| \leq E, & k \in K \end{aligned}$$

Result: REPR has Lower MSE Variance across Folds

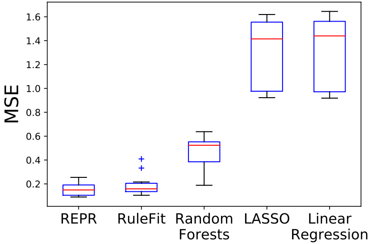
Data: HEAT



Data: MACHINE



Data: SERVO



Data: CONCRETE