Optimal accuracy threshold using 0-1 ILP formulation

Dr. Pouria Ramazi, Arash Mari Oriyad

Abstract

Classification

For a binary classification problem, the accuracy measure which is the portion of the correctly predicted instances with respect to all instances, can be calculated as below,

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP, and FN are the number of true-positive, true-negative, false-positive, and false-negative instances, respectively.

While accuracy requires the predicted labels, classification models produce a probability distribution over all available classes for each instance. Hence, one must apply a threshold to the predicted probabilities to obtain deterministic labels. An optimal threshold for a given set of instances is a real value between 0 and 1, providing the highest accuracy.

A trivial approach to determine such a threshold is an exhaustive search that examines all the possible threshold values to obtain the best one. Complete search-based methods are usually not efficient in execution time, especially when the solution space grows. To the best of our knowledge, there is no programming package to find the optimal accuracy threshold without using an exhaustive search.

We treated the optimal accuracy threshold problem as an optimization problem by introducing an integer linear programming (ILP) formulation. Assume a set of N instances with a binary target variable y and suppose $\hat{y}_i$ is the predicted probability for the ith instance. Then, $\tilde{y}_i$ is a binary decision variable which is desirable to be calculated as follows:

$$\tilde{y}_i = \begin{cases} 0, & \hat{y}_i < t \\ 1, & \hat{y}_i \geq t \end{cases}$$

where t is the threshold. More precisely, $\tilde{y}_i$ represents the label of ith instance after applying the threshold t to the predicted probability $\hat{y}_i$. Hence, the accuracy measure can be calculated as bellow.

$$accuracy = 1 - \frac{1}{N}\sum_{i=1}^{N} |y_i - \tilde{y}_i|$$

Put them all together, the following optimization formulation is achieved.

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{N} |y_i - \tilde{y}_i| \\
\text{subject to:} \quad & \tilde{y}_i = \begin{cases} 0, & \hat{y}_i < t \\ 1, & \hat{y}_i \geq t \end{cases} \qquad \forall 1 \leq i \leq N \\
& \tilde{y}_i \in \{0, 1\} \qquad \forall 1 \leq i \leq N
\end{aligned} \tag{1}
$$

Note that y_i and $\hat{y}_i$ are input parameters and $\tilde{y}_i$ is the decision variable of the problem. Moreover, $\frac{1}{N}$ is ignored in the objective function since N does not affect the final solution.

To transform the proposed constraints into a linear form, for each i, we can combine the conditions and replace them with two new constraints. Similarly, to make the objective function linear, a new binary decision variable $z_i = |y_i - \tilde{y}_i|$ is defined and two new constraints are added to the optimization formulation as follows:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{N} z_i \\
\text{subject to:} \quad & z_i \geq y_i - \tilde{y}_i && \forall 1 \leq i \leq N \\
& z_i \geq \tilde{y}_i - y_i && \forall 1 \leq i \leq N \\
& \hat{y}_i (1 - \tilde{y}_i) < t && \forall 1 \leq i \leq N \\
& \tilde{y}_i (\hat{y}_i - 1) \geq t - 1 && \forall 1 \leq i \leq N \\
& \tilde{y}_i \in \{0, 1\} && \forall 1 \leq i \leq N \\
& 0 \leq t \leq 1
\end{aligned} \tag{2}
$$

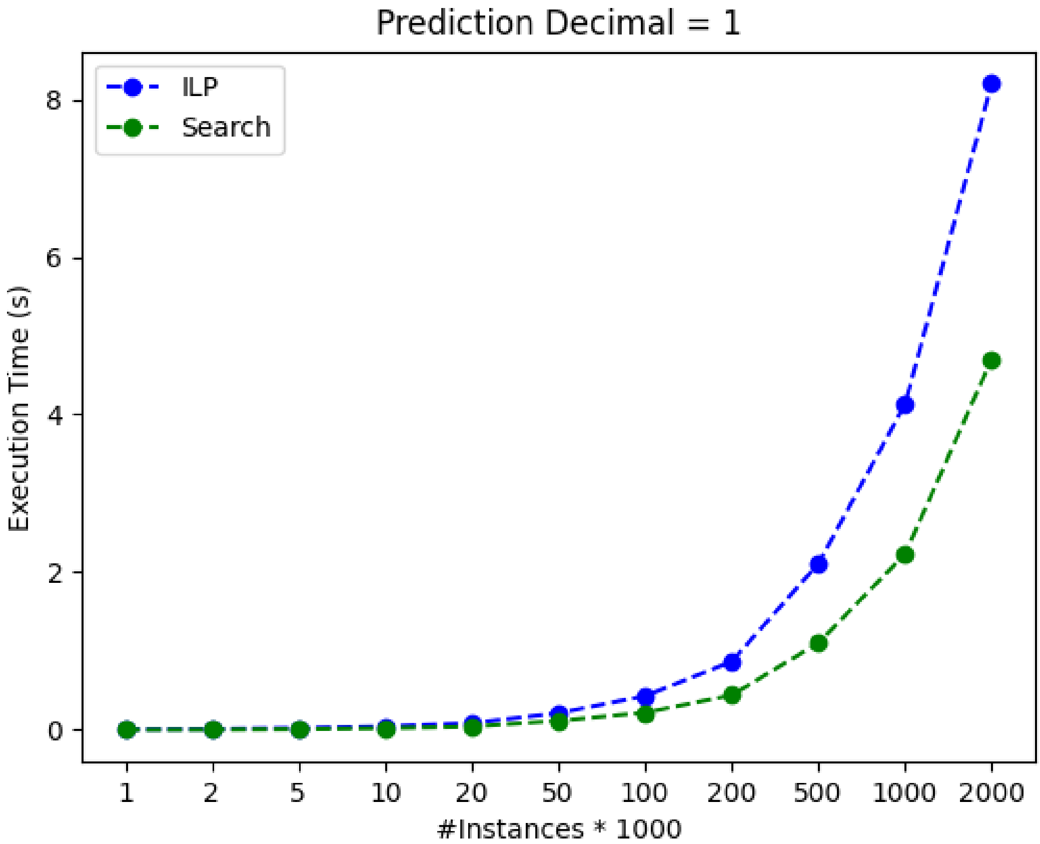
Hence, we have designed a 0-1 ILP formulation with $2N$ binary decision variables and $4N$ linear constraints.

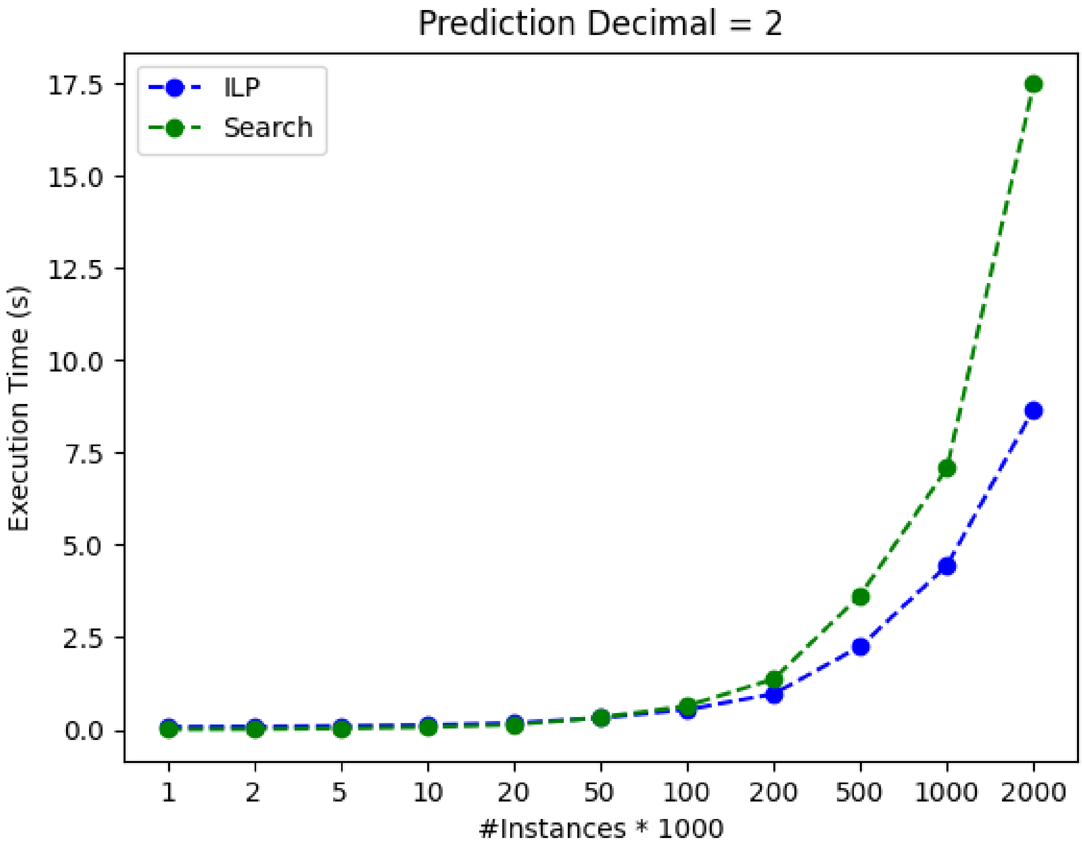
The proposed ILP formulation is compared with the exhaustive search method in terms of average execution time (Table 1).

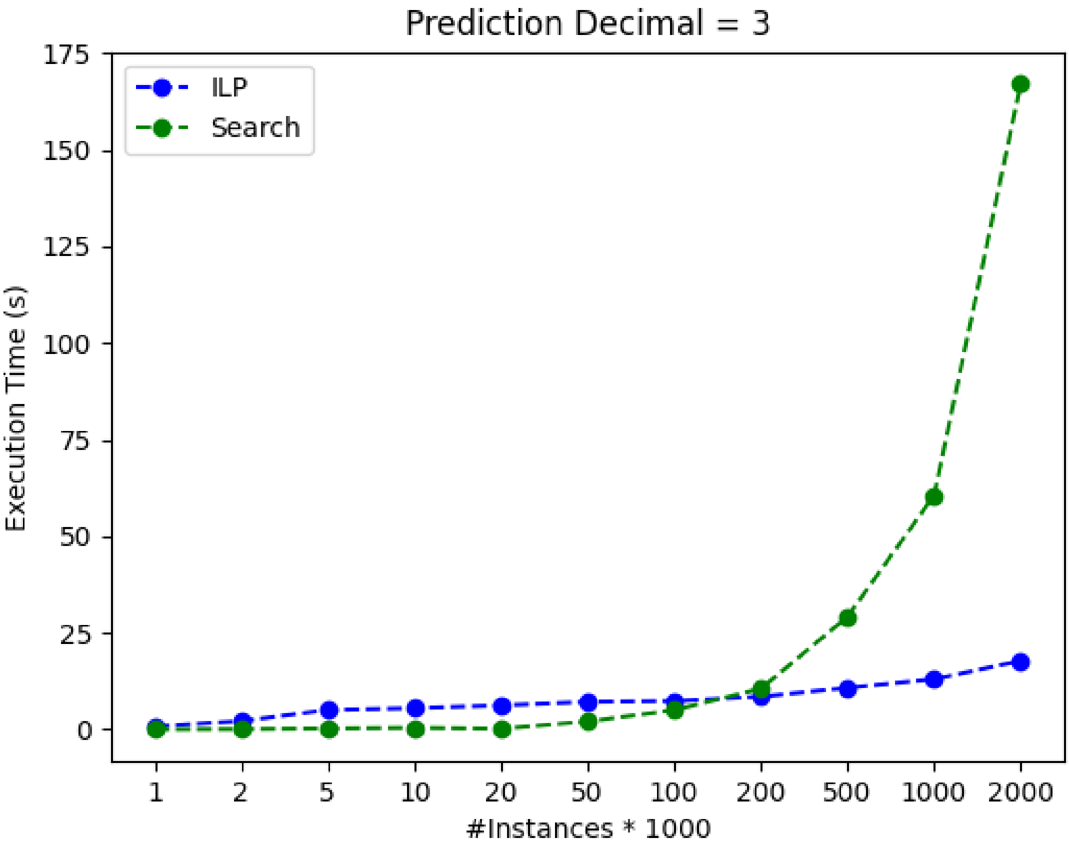
Table 1: Comparing the ILP-based approach with the exhaustive search method for finding the optimal accuracy threshold.

# Instances	Prediction Decimal	ILP (Python)	Exhaustive Search (Python)	ILP (C++)	Exhaustive Search (C++)
1000	1	0.032	0.025	0.0068	0.0024
1000	2	0.150	0.199	0.067	0.0062
1000	3	1.133	1.243	0.802	0.028
1000	4	1.748	1.884	1.311	0.041
1000	5	1.807	1.860	1.339	0.045
2000	1	0.033	0.026	0.010	0.0042
2000	2	0.160	0.205	0.076	0.013
2000	3	2.747	1.721	2.152	0.076
2000	4	6.378	3.533	5.173	0.159
2000	5	?	?	?	0.172
5000	1	?	?	0.023	0.010
5000	2	?	?	0.098	0.030
5000	3	5.998	2.075	5.060	0.234
5000	4	?	?	?	0.875
5000	5	50.663	9.747	?	1.091
10^4	1	?	?	0.043	0.0216
10^4	2	?	?	0.127	0.0634
10^4	3	?	?	5.527	0.397
2×10^4	1	?	?	0.086	0.0421
2×10^4	2	?	?	0.179	0.130
2×10^4	3	?	?	6.227	0.818
5×10^4	1	?	?	0.212	0.115

5×10^4	2	?	?	0.314	0.330
5×10^4	3	?	?	7.115	2.059
10^5	1	?	?	0.429	0.218
10^5	2	?	?	0.539	0.642
10^5	3	?	?	7.351	5.014
2×10^5	1	?	?	0.866	0.439
2×10^5	2	?	?	0.966	1.370
2×10^5	3	?	?	8.462	10.518
5×10^5	1	?	?	2.108	1.102
5×10^5	2	?	?	2.254	3.642
5×10^5	3	?	?	10.771	29.151
10^6	1	?	?	4.131	2.229
10^6	2	?	?	4.451	7.088
10^6	3	?	?	13.010	60.555
2×10^6	1	?	?	8.198	4.703
2×10^6	2	?	?	8.692	17.500
2×10^6	3	?	?	17.751	166.988


Prediction Decimal = 1
ILP
Search
Execution Time (s)
0
2
4
6
8
1
2
5
10
20
50
100
200
500
1000
2000
#Instances * 1000


Prediction Decimal = 2
ILP
Search
Execution Time (s)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
1
2
5
10
20
50
100
200
500
1000
2000
#Instances * 1000


Prediction Decimal = 3
ILP
Search
Execution Time (s)
#Instances * 1000
0
25
50
75
100
125
150
175
1
2
5
10
20
50
100
200
500
1000
2000

For $N = 2$:

$$
\begin{aligned}
\text{minimize} \quad & z_1 + z_2 \\
\text{subject to:} \quad & z_1 + \tilde{y}_1 \geq y_1 \\
& z_2 + \tilde{y}_2 \geq y_2 \\
& z_1 - \tilde{y}_1 \geq -y_1 \\
& z_2 - \tilde{y}_2 \geq -y_2 \\
& \hat{y}_1 \tilde{y}_1 + t \geq \hat{y}_1 + \epsilon \\
& \hat{y}_2 \tilde{y}_2 + t \geq \hat{y}_2 + \epsilon \\
& (\hat{y}_1 - 1)\tilde{y}_1 - t \geq -1 \\
& (\hat{y}_2 - 1)\tilde{y}_2 - t \geq -1
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
\text{minimize} \quad & wx \\
\text{subject to:} \quad & Ax \geq b
\end{aligned}
\tag{4}
$$

$$
A = \begin{array}{c}
\begin{array}{ccccc} z_1 & z_2 & \tilde{y}_1 & \tilde{y}_2 & t \end{array} \\
\left(\begin{array}{cc|cc|c}
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 \\
\hline
1 & 0 & -1 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 \\
\hline
0 & 0 & \hat{y}_1 & 0 & 1 \\
0 & 0 & 0 & \hat{y}_2 & 1 \\
\hline
0 & 0 & \hat{y}_1 - 1 & 0 & -1 \\
0 & 0 & 0 & \hat{y}_2 - 1 & -1
\end{array}\right)_{8\times 5}
\end{array}
$$

$$
x = \left(\begin{array}{c}
z_1 \\
z_2 \\
\hline
\tilde{y}_1 \\
\tilde{y}_2 \\
\hline
t
\end{array}\right)_{5\times 1}
$$

$$
b = \left(\begin{array}{c}
y_1 \\
y_2 \\
\hline
-y_1 \\
-y_2 \\
\hline
\hat{y}_1 + \epsilon \\
\hat{y}_2 + \epsilon \\
\hline
-1 \\
-1
\end{array}\right)_{8\times 1}
$$

$$A = \begin{array}{c} \begin{array}{ccccccccc} z_1 & z_2 & \dots & z_N & \tilde{y}_1 & \tilde{y}_2 & \dots & \tilde{y_N} & t \end{array} \\ \left(\begin{array}{cccc|cccc|c}
1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \\
0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 & 0 \\
\hline
1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 & 0 \\
0 & 1 & \dots & 0 & 0 & -1 & \dots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \dots & 1 & 0 & 0 & \dots & -1 & 0 \\
\hline
0 & 0 & \dots & 0 & \hat{y}_1 & 0 & 0 & 0 & 1 \\
0 & 0 & \dots & 0 & 0 & \hat{y}_2 & 0 & 0 & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \dots & 0 & 0 & 0 & 0 & \hat{y_N} & 1 \\
\hline
0 & 0 & \dots & 0 & \hat{y}_1 - 1 & 0 & 0 & 0 & -1 \\
0 & 0 & \dots & 0 & 0 & \hat{y}_2 - 1 & 0 & 0 & -1 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \dots & 0 & 0 & 0 & 0 & \hat{y_N} - 1 & -1
\end{array}\right)_{4N\times(2N+1)} \end{array}$$

$$recall = \frac{TP}{TP+FN} = \frac{TP}{\#\ \text{class } 1}$$

$$\begin{aligned}
\text{maximize} \quad & \sum_{i=1}^{N} y_i \tilde{y}_i \\
& \hat{y}_i(1-\tilde{y}_i) < t && \forall 1 \le i \le N \\
& \tilde{y}_i(\hat{y}_i - 1) \ge t-1 && \forall 1 \le i \le N \\
& \tilde{y}_i \in \{0,1\} && \forall 1 \le i \le N \\
& 0 \le t \le 1
\end{aligned} \tag{5}$$

$$\max \quad \alpha \times \text{accuracy} + \beta \times \text{recall}$$

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{N} \alpha z_i - \beta y_i \tilde{y}_i \\
\text{subject to:} \quad & z_i \geq y_i - \tilde{y}_i & \forall 1 \leq i \leq N \\
& z_i \geq \tilde{y}_i - y_i & \forall 1 \leq i \leq N \\
& \hat{y}_i (1 - \tilde{y}_i) < t & \forall 1 \leq i \leq N \\
& \tilde{y}_i (\hat{y}_i - 1) \geq t - 1 & \forall 1 \leq i \leq N \\
& \tilde{y}_i \in \{0, 1\} & \forall 1 \leq i \leq N \\
& 0 \leq t \leq 1
\end{aligned}
\tag{6}
$$

$$precision = \frac{TP}{TP + FP} = \frac{TP}{\text{\# class 1}}$$

$$\max \quad \frac{\sum y_i \tilde{y}_i}{\sum \tilde{y}_i}$$