CASE STUDY: Classification by Maximizing Area Under ROC Curve (AUC)
(pr_pen (0, L3), pr_pen(0, L1-L2))

Background

In classification problems, AUC (Area Under ROC Curve) is a popular measure for evaluating the goodness of classifiers. Namely, a classifier which attains higher AUC is preferable to a lower AUC classifier. This motivates us directly maximize AUC for obtaining a classifier. Such a direct maximization is rarely applied because of numerical difficulties. Usually, computationally tractable optimization methods (e.g., logistic regression) are used, and then, their goodness is measured by AUC. This case study maximizes the AUC in a direct manner. Similar approach is used in paper [1], which considers an approximation approach to AUC maximization. More precisely, [1] maximizes a surrogate of AUC by replacing the 0-1 step function, which appears in a representation of AUC, with a sigmoid function, so that the derivative can be calculated. In contrast with [1], this case study directly maximizes AUC. Difficulties of the associated maximization problem are: (a) Objective function is nonconvex; (b) Number of scenarios may be huge so that standard PCs cannot handle all the scenarios at once in operation memory. We present AUC as follows:

AUC = 1- probability ((difference of two linear functions with random coefficients) <=0).

PSG can calculate and minimize the probability of the difference of two linear functions with independent random vectors of coefficients. This case study maximizes AUC by minimizing PSG probability function `pr_pen`. Two equivalent variants are considered: 1)Problem 1 uses difference of two independent random linear functions presented by two different matrices of scenarios (with the same column headers); 2) Problem 2 uses one matrix of scenarios which is manually generated by taking differences of linear functions from two different matrices (this matrix can be created only for small dimensions because the number of rows in the resulting matrix equals the product of number of rows in the first and in the second matrix). Problems 1 and 2 are mathematically equivalent, but they have different data inputs.

Problem 3 solves logistic regression problem by maximizing PSG `logexp_sum` function using Matrix of Scenarios with binary "benchmark" (dependent variable). Further, this Matrix of Scenarios is divided in two different Matrices of Scenarios. The first Matrix of Scenarios includes rows with 1 benchmark value, and the second Matrix of Scenarios includes rows with 0 benchmark value. These two created matrices have zero benchmark values. Then, we evaluate AUC of this logistic regression classifier by calculating probability (`pr_pen`) of difference of two random linear functions presented by two different matrices of scenarios. Optimal logistic regression point is used as an initial point in the following Problems 4 and 5.

Problem 4 maximizes AUC by minimizing PSG probability function `pr_pen` , using difference of two random linear functions presented by two different matrices of scenarios under a linear constraint. This Problem uses the optimal logistic regression point (obtained in the Problem 3) as an initial point.

Problem 5 maximizes AUC by minimizing PSG probability function `pr_pen` , using one matrix of scenarios which is generated by taking differences of linear functions from two different matrices. The problem is solved with a linear constraint. This Problem uses the optimal logistic regression point (obtained in the Problem 3) as an initial point.

Problems 3-5 are solved for 3 datasets generated from one large data set.

References

[1] Miura K, Yamashita S, Eguchi S. (2010) Area Under the Curve Maximization Method in Credit Scoring.The Journal of Risk Model Validation, 4(2), pp.3-25.

Notations

$\vec{x} = (x_0, x_1, \dots, x_I)$ = vector of decision variables;

$\vec{\theta}^1 = (\theta_0^1, \theta_1^1, \dots, \theta_I^1)$ = first random scenario vector;

$\vec{\theta}^1_{j_1} = \left(\theta^1_{j_1 0}, \theta^1_{j_1 1}, \dots, \theta^1_{j_1 I}\right)$ = scenario j_1 of the first random scenario vector;

$L\left(\vec{x}, \vec{\theta}^1\right) = \theta^1_0 - \sum_{i=1}^{I} \theta^1_i x_i$ = linear loss function with random vector $\vec{\theta}^1$;

$L\left(\vec{x}, \vec{\theta}^1_{j_1}\right)$ = scenario j_1 of random loss function $L\left(\vec{x}, \vec{\theta}^1\right)$, $j_1 = 1, \dots, J_1$;

$p^1_{j_1}$ = probability of scenario $\left(\vec{x}, \vec{\theta}^1_{j_1}\right)$, $j_1 = 1, \dots, J_1$;

$\vec{\theta}^2 = (\theta^2_0, \theta^2_1, \dots, \theta^2_I)$ = second random scenario vector;

$\vec{\theta}^2_{j_2} = \left(\theta^2_{j_2 0}, \theta^2_{j_2 1}, \dots, \theta^2_{j_2 I}\right)$ = scenario j_2 of the second random scenario vector;

$L\left(\vec{x}, \vec{\theta}^2\right) = \theta^2_0 - \sum_{i=1}^{I} \theta^2_i x_i$ = second linear loss function with random vector $\vec{\theta}^2$;

$L\left(\vec{x}, \vec{\theta}^2_{j_2}\right)$ = scenario j_2 of random loss function $L\left(\vec{x}, \vec{\theta}^2\right)$, $j_2 = 1, \dots, J_2$;

$p^2_{j_2}$ = probability of scenario $\left(\vec{x}, \vec{\theta}^2_{j_2}\right)$, $j_2 = 1, \dots, J_2$;

$L\left(\vec{x}, \vec{\theta}^3\right)$ = $\theta^3_0 - \sum_{i=1}^{I} \theta^3_i x_i$ = third linear loss with random vector $\vec{\theta}^3$ = $\vec{\theta}^1 - \vec{\theta}^2$;

$\vec{\theta}^3_{j_3} = \left(\theta^3_{j_3 0}, \theta^3_{j_3 1}, \dots, \theta^3_{j_3 I}\right)$ = scenario j_3 of the third random scenario vector, row $j_3 = j_1 \cdot j_2$ of this matrix equals the difference $\vec{\theta}^3_{j_3} = \vec{\theta}^1_{j_1} - \vec{\theta}^2_{j_2}$, $j_3 = 1, \dots, J_1 \cdot J_2$;

Because of linearity of loss functions, scenario j_3 of the third linear function $L\left(\vec{x}, \vec{\theta}^3\right)$ equals $L\left(\vec{x}, \vec{\theta}^3_{j_3}\right)$= $L\left(\vec{x}, \vec{\theta}^1_{j_1}\right) - L\left(\vec{x}, \vec{\theta}^2_{j_2}\right)$, $j_3 = 1, \dots, J_1 \cdot J_2$;

$p^3_{j_3}$ = $p^1_{j_1} \cdot p^2_{j_2}$= probability of scenario j_3 of loss function $L\left(\vec{x}, \vec{\theta}^3\right)$, $j_3 = 1, \dots, J_1 \cdot J_2$;

$pr\_pen\left(w, L\left(\vec{x}, \vec{\theta}\right)\right)$=$\Pr\left\{L\left(\vec{x}, \vec{\theta}\right) \geq w\right\}$ = PSG function Probability Exceeding Penalty for Loss based on scenario matrix for random vector $\vec{\theta}$:

$pr\_pen\left(w, L\left(\vec{x}, \vec{\theta}\right)\right) = \sum_{j=1}^{J} p_j h\left(w, L\left(\vec{x}, \vec{\theta}\right)\right)$,

where $h(w, y) = \begin{cases} 1, & \text{if } y \geq w, \\ 0, & \text{otherwise} \end{cases}$;

$pr\_pen\left(w, L^1\left(\vec{x}, \vec{\theta}^1\right) - L^2\left(\vec{x}, \vec{\theta}^2\right)\right)$= PSG function Probability Exceeding Penalty for Loss applied to the difference of linear loss function $L^1\left(\vec{x}, \vec{\theta}^1\right) - L^2\left(\vec{x}, \vec{\theta}^2\right)$, this PSG function, as input, has two scenario matrices for the random vectors $\vec{\theta}^1$ and $\vec{\theta}^2$;

Log-likelihood function in logistic regression presented by PSG function Logarithms Exponents Sum `(logexp_sum)`:

$$logexp\_sum\left(\vec{x}, \vec{\theta}\right) = \sum_{j=1}^{J} p_j ln\left[\frac{c_j exp\{\sum_{i=1}^{I} \theta_{ij} x_i\}}{c_j exp\{\sum_{i=1}^{I} \theta_{ij} x_i\} + (1 - c_j) c_j exp\{-\sum_{i=1}^{I} \theta_{ij} x_i\}}\right],\ 0 < c_j < 1. \ .$$

Optimization Problem 1

minimizing probability that loss function $L\left(\vec{x}, \vec{\theta}^3\right)$ with $J_1 \cdot J_2$scenarios is below zero

$$\min_{\vec{x}}\ pr\_pen\left(0, L\left(\vec{x}, \vec{\theta}^3\right)\right) \tag{CS.1}$$

Optimization Problem 2

minimizing probability that difference of two loss functions $L^1(\vec{x},\vec{\theta}^1) - L^2(\vec{x},\vec{\theta}^2)$ *with* J_1 *and* J_2 *scenarios respectively is below zero*

$$\min_{\vec{x}} \; pr\_pen\left(0, L(\vec{x},\vec{\theta}^1) - L(\vec{x},\vec{\theta}^2)\right) \tag{CS.2}$$

Optimization Problem 3

maximizing log-likelihood function for logistic regression

$$\max_{\vec{x}} \; logexp\_sum(\vec{x},\vec{\theta}) \tag{CS.3}$$

$$Calculate\left(pr\_pen\left(0, L(\vec{x}^*,\vec{\theta}^1) - L(\vec{x}^*,\vec{\theta}^2)\right)\right) \text{, where } \vec{x}^* = \arg\max_{\vec{x}}\left(logexp\_sum(\vec{x},\vec{\theta})\right)$$

Optimization Problem 4

minimizing probability that difference of two loss functions $L(\vec{x},\vec{\theta}^1) - L(\vec{x},\vec{\theta}^2)$ *with* J_1 *and* J_2 *scenarios respectively is below zero*

$$\min_{\vec{x}} \; pr\_pen\left(0, L(\vec{x},\vec{\theta}^1) - L(\vec{x},\vec{\theta}^2)\right) \tag{CS.6}$$

subject to

linear constraint

$$\sum_{i=1}^{I} a_i\, x_i = b \tag{CS.7}$$

$$\text{initial point} = \vec{x}^* = \arg\max_{\vec{x}} \left(logexp\_sum(\vec{x},\vec{\theta})\right) \tag{CS.8}$$

Optimization Problem 5

minimizing probability that loss function $L(\vec{x},\vec{\theta}^3)$ *with* $J_1 \cdot J_2$ *scenarios is below zero*

$$\min_{\vec{x}} \; pr\_pen(0, L(\vec{x},\vec{\theta}^3)\,) \tag{CS.9}$$

subject to

$$\sum_{i=1}^{I} a_i\, x_i = b \tag{CS.10}$$

$$\text{initial point} = \vec{x}^* = \arg\max_{\vec{x}} \left(logexp\_sum(\vec{x},\vec{\theta})\right) \tag{CS.11}$$