Method for computing muticlass ROC AUC

Dataiku DSS
www.dataiku.com

ABSTRACT

This document explains the methodology implemented to compute the ROC AUC for multiclass classification models in Dataiku DSS.

General Formula

We define MROC_{AUC} as an equivalent of ROC_{AUC} for multiclass classification.
Let C be the number of classes,

$$\text{MROC}_{\text{AUC}} = \frac{1}{C \times (C-1)} \cdot \sum_{i=0}^{C-1} \sum_{j=0, j \neq i}^{C-1} A(i,j)$$

Detail of how $A(i,j)$ is computed

Input arrays

Let y_{truth} be the array of ground truth class values (in $\{0, ..., C-1\}$). It is of shape $(M, 1)$, i.e. M rows and 1 column:

$$y_{\text{truth}} = \begin{pmatrix} 3 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \left.\right\} \begin{array}{c} M \text{ rows,} \\ \text{values} \in \{0, ..., C-1\} \end{array}$$

Let y_{probas} be the array of predicted probabilities. Each column i corresponds to a class i, where values are probability estimates for the class i. It is of shape (M, C), i.e. M rows and C columns:

$$y_{\text{probas}} = \begin{pmatrix} 0.1 & 0.9 & ... & 0 \\ 0.8 & 0 & ... & 0.2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & 0.7 \\ 0 & 0.1 & ... & 0.9 \end{pmatrix} \left.\right\} M \text{ rows}$$

$$\underbrace{}_{C \text{ cols}}$$

$A(i,j)$ computation

For every pair of classes $i \neq j$, let $L_{i,j}$ be the subset of rows of y_{truth} where the ground truth is either i or j.

$$L_{i,j} = \left\{ k \in \{0, ..., M-1\} \mid y_{\text{truth},k} = i \text{ or } y_{\text{truth},k} = j \right\}$$

With $y_{\text{truth},L_{i,j}}$ and $y_{\text{proba},L_{i,j}}$ the corresponding arrays with only these rows
- Let $y_{\text{truth},L_{i,j},\text{binarized}}$ be a copy of $y_{\text{truth},L_{i,j}}$ with all values at j set to 0 and all at i set to 1
- Note $y_{\text{probas},L_{i,j}, \text{col } i}$ the column i of $y_{\text{probas},L_{i,j}}$

$$A(i,j) = \text{ROC}_{\text{AUC}}\left(y_{\text{truth},L_{i,j},\text{binarized}}, y_{\text{probas},L_{i,j}, \text{col } i} \right)$$

with ROC_{AUC} the usual binary classification ROC AUC metric.