AUCμ: A Performance Metric for Multi-Class Machine Learning Models

Ross Kleiman & David Page

University of Wisconsin – Madison

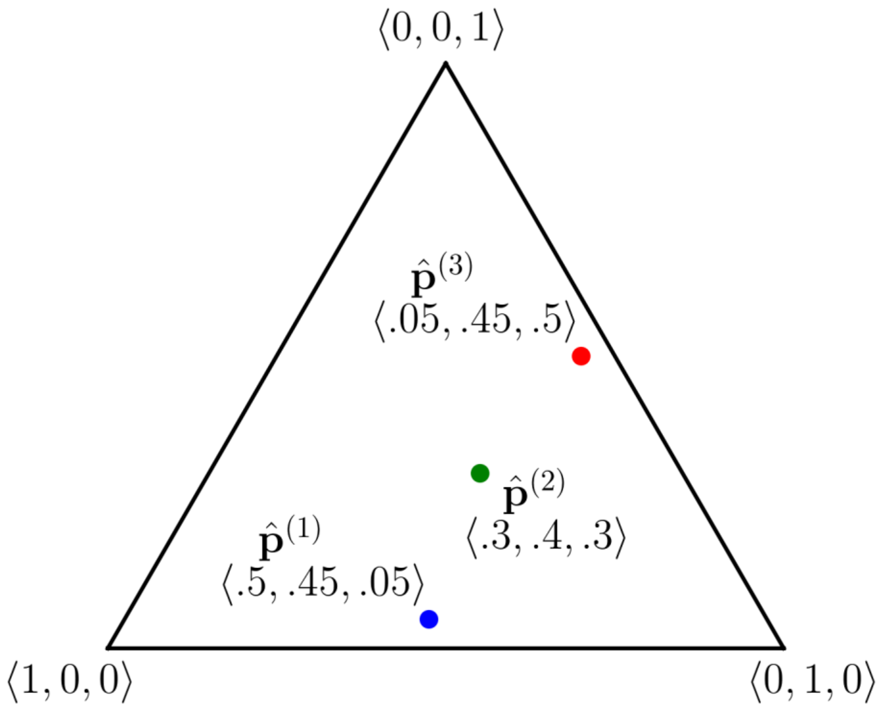
Problem Setting

Motivation

- Predicting 1 of K classes is a multi-class problem
- AUC cannot be used to evaluate multiclass models

Background

- Soft-classifications live on the $K-1$ simplex, Δ_{K-1}



Multi-Class AUC Survey

Composition approaches:

- Hand & Till 2001
- Provost & Domingos 2000
  - Class-skew sensitive
- Both H&T and P&D can underscore models

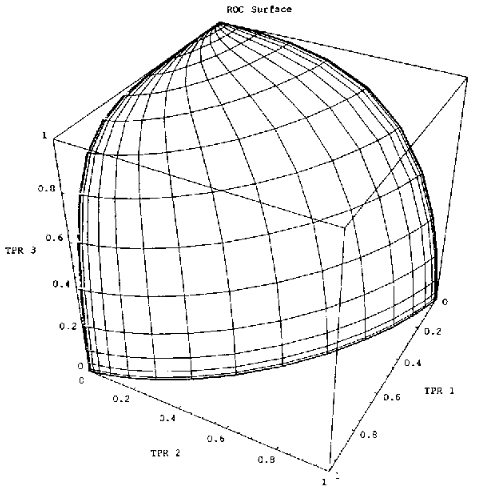
Multi-Class AUC Survey

Composition approaches:

- Hand & Till 2001
- Provost & Domingos 2000
    - Class-skew sensitive
- Both H&T and P&D can underscore models

Volume under the ROC surface (VUS):

- Mossman 1999 (3-way ROCs)
- Ferri, *et al.* 2003
- Drawbacks
    - Computational complexity
    - Don’t range from 0-1
    - Random guessing not 0.5



Mossman 1999

Motivating Properties

Key Properties of two-class AUC

1. If a model gives the correct label to the highest probability on each example, then AUC=1
2. Random guessing on examples yields AUC=0.5
3. AUC is insensitive to class skew

Motivating Properties

Key Properties of two-class AUC

1. If a model gives the correct label to the highest probability on each example, then AUC=1
2. Random guessing on examples yields AUC=0.5
3. AUC is insensitive to class skew

These can all be derived from the Mann-Whitney U-Statistic

$$U = \frac{1}{n_+ n_-} \sum_{\hat{p}^{(a)} \in D(+)} \sum_{\hat{p}^{(b)} \in D(-)} \tilde{I}\left(\hat{p}^{(a)} - \hat{p}^{(b)}\right)$$

$$\tilde{I} = \begin{cases} 1 \text{ if } \hat{p}^{(a)} > \hat{p}^{(b)} \\ 0 \text{ if } \hat{p}^{(a)} < \hat{p}^{(b)} \\ 0.5 \text{ if } \hat{p}^{(a)} = \hat{p}^{(b)} \end{cases}$$

Definition of AUC$_\mu$

We extend the Mann-Whitney U-statistic to the multi-class setting

$$AUC_\mu = \frac{2}{K(K-1)} \sum_{i<j} \frac{1}{n_i n_j} \sum_{a \in D(i), b \in D(j)} \tilde{I}\left(O\left(y^{(a)}, y^{(b)}, \hat{p}^{(a)}, \hat{p}^{(b)}, v_{i,j}\right)\right)$$

Note that $\tilde{I}$ is the modified indicator function

O is the orientation function for two predictions

$$O\left(y^{(a)}, y^{(b)}, \hat{p}^{(a)}, \hat{p}^{(b)}, v_{i,j}\right) = \left(v_{i,j} \cdot \left(y^{(a)} - y^{(b)}\right)\right)\left(v_{i,j} \cdot \left(\hat{p}^{(a)} - \hat{p}^{(b)}\right)\right)$$

Analysis of AUC_μ

	VUS-3	VUS	H&T	P&D	AUC_μ
Perfect = 1	$\surd$	$\times$	$\times$	$\times$	$\surd$
Random = 0.5	$\times$	$\times$	$\surd$	$\surd$	$\surd$
Skew Insensitive	?	?	$\surd$	$\times$	$\surd$
Time Complexity	exponential	exponential	polynomial	polynomial	polynomial

Time Complexity: $O(Kn \log n)$ with *argmax* partition matrix

$O\big(Kn(K + \log n)\big)$ with an arbitrary partition matrix

Many Thanks To

- Page Research Group
  - **David Page (author/advisor)**
  - Wei Zhang (discussion and review)
  - Charles Kuang (discussion and review)
- Funding Sources
  - CIBM training grant: 5T15LM007359
  - NCATS grant: UL1TR002373