UNIVERSITY OF COPENHAGEN

Funded by the European Union

AIPHY

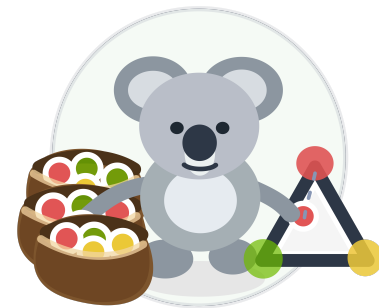
Multiclass Classification without Labels via Posterior Simplex Geometry

Date: 19 August 2026

Venue: HAMLET-PHYSICS 2026

By: Janni — Niels Bohr Institute, University of Copenhagen

Authors: Raphaël Bonnet-Guerrini, Johann Ioannou-Nikolaides, Troels Petersen, Vincenzo Piuri



The problem

We have an unlabeled dataset, no per-example ground truth, and no way to get one cheaply

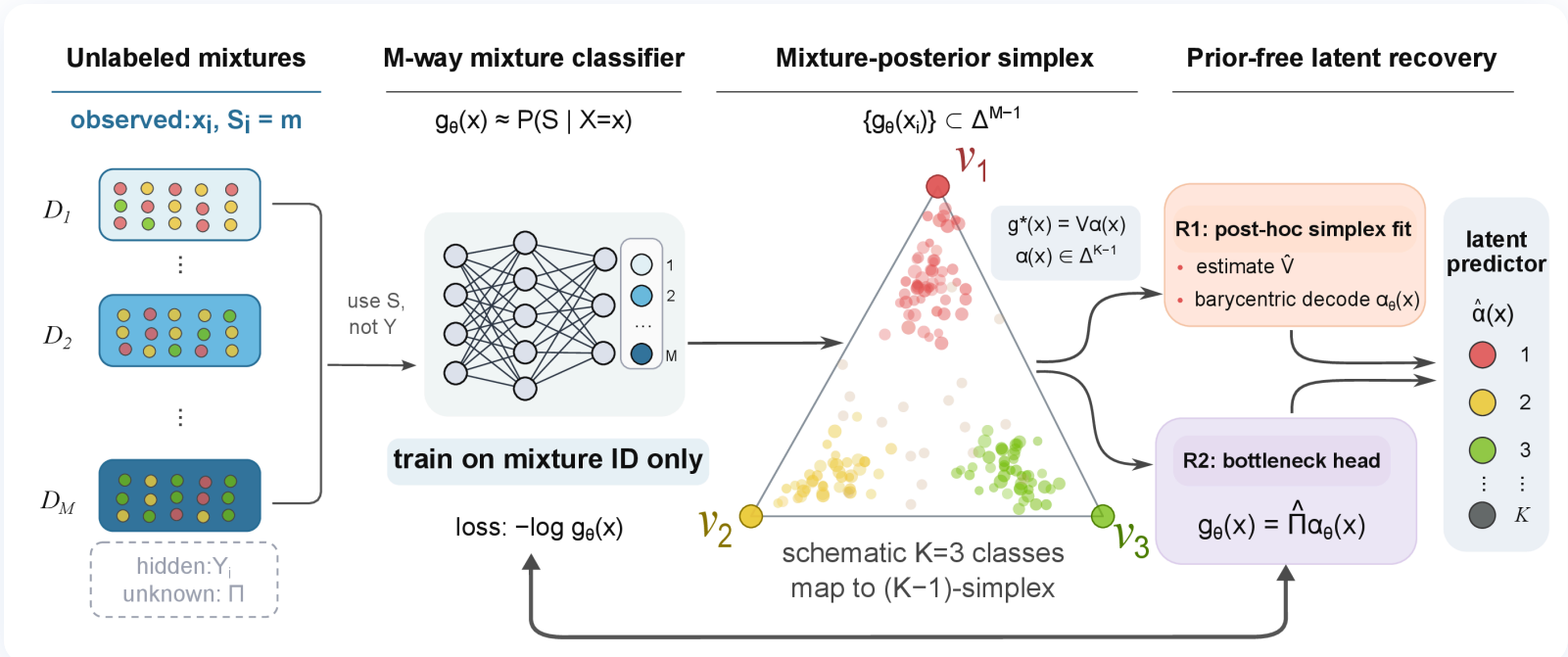
- But we *can* split it into several **mixtures** with different, even unknown, class proportions: different cuts, triggers, detectors, or observing conditions
- *Classification without labels: Learning from mixed samples in high energy physics* (CWoLa) showed this is enough to solve the problem for two classes ($K = 2$)

What if we have more than two classes?

The idea

Train an ordinary classifier to predict **only which mixture** each example came from

Posterior probabilities organize onto a simplex, the K corners of that simplex are exactly the hidden classes



When is this useful?

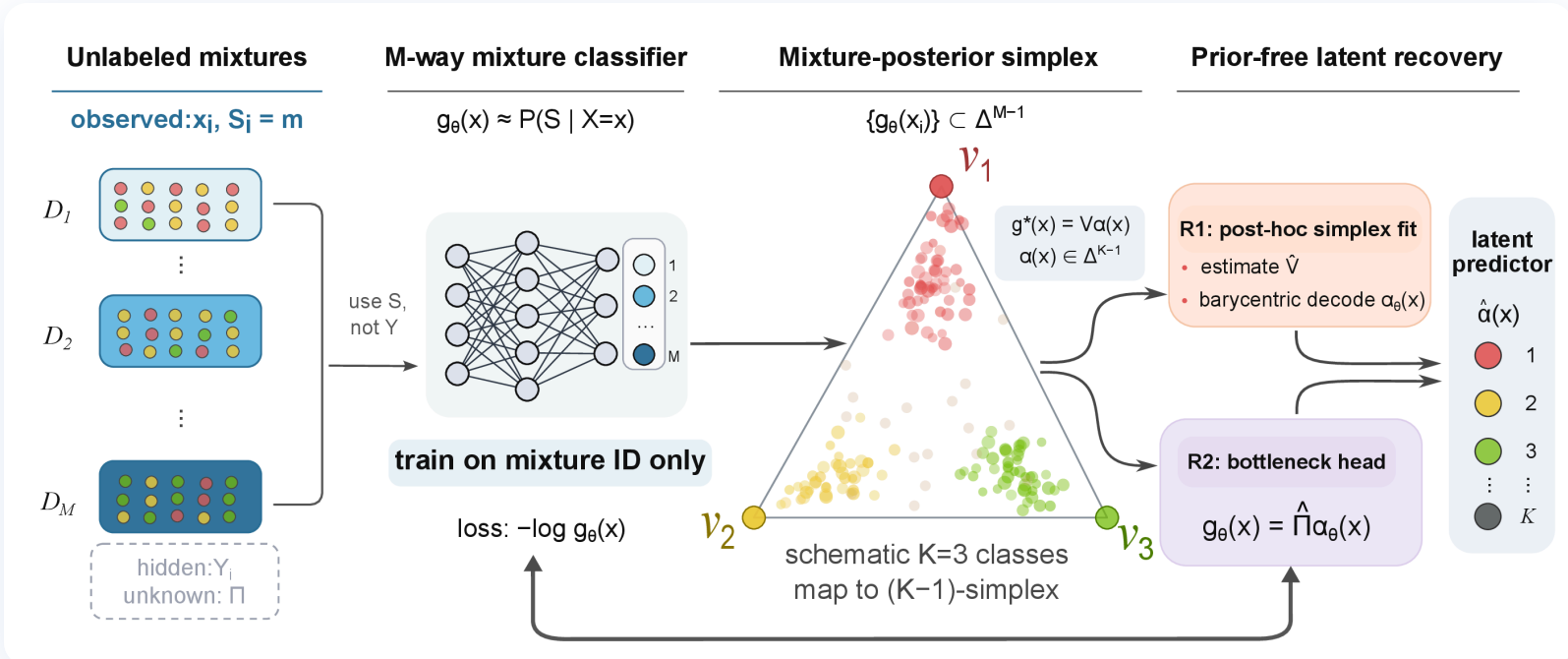
Use it when

You have, or can construct, several mixed, unlabeled datasets with different (even unknown) class fractions, but no reliable ground truth / simulation

You get

Calibrated per-example class probabilities, *and* the mixing matrix Π: the class fractions inside every mixture

Summary


Unlabeled mixtures
observed:x_i, S_i = m
D_1
D_2
D_M
hidden:Y_i
unknown: Π
use S, not Y
M-way mixture classifier
$g_\theta(x) \approx P(S \mid X=x)$
1
2
M
train on mixture ID only
loss: −log $g_\theta(x)$
Mixture-posterior simplex
$\{g_\theta(x_i)\} \subset \Delta^{M-1}$
v_1
v_2
v_3
schematic K=3 classes
map to (K−1)-simplex
g*(x) = Vα(x)
α(x) ∈ Δ^{K-1}
Prior-free latent recovery
R1: post-hoc simplex fit
• estimate $\hat{V}$
• barycentric decode $\alpha_\theta(x)$
R2: bottleneck head
$g_\theta(x) = \hat{\Pi}\alpha_\theta(x)$
latent predictor
$\hat{\alpha}(x)$
1
2
3
K

Thank You

Multiclass Classification without Labels via Posterior Simplex Geometry

Johann Ioannou-Nikolaides

Niels Bohr Institute, University of Copenhagen

johann.nikolaides@nbi.ku.dk

arXiv:2607.24943

Come find the poster for the full story:
assumptions, fitters, and all the numbers

Questions?