An Embedding Framework for Calibrated Polyhedral Surrogates

Jessie Finocchiaro, Rafael Frongillo, Bo Waggoner

University of Colorado Boulder

Introduction

In classification and classification-like problems, we often design discrete loss functions that are intuitive and match the task at hand (e.g. 0-1 loss captures the essence of multiclass classification.) These discrete losses are hard to optimize, so we design **surrogate loss functions** to model our original, discrete loss. Often, surrogates are designed in an ad hoc manner, and their calibration is not guaranteed. *We show how to construct calibrated, polyhedral surrogates for such problems*.

Setting

$\ell : \mathcal{R} \to \mathbb{R}^n_+$ — Discrete loss

$L : \mathbb{R}^d \to \mathbb{R}^n_+$ — Surrogate loss

$p \in \Delta_{\mathcal{Y}}: \langle p, L(u) \rangle$ — Expected loss

$\underline{L} : p \mapsto \inf_{u \in \mathbb{R}^d} \langle p, L(u) \rangle$ — Bayes Risk

Embedding Framework

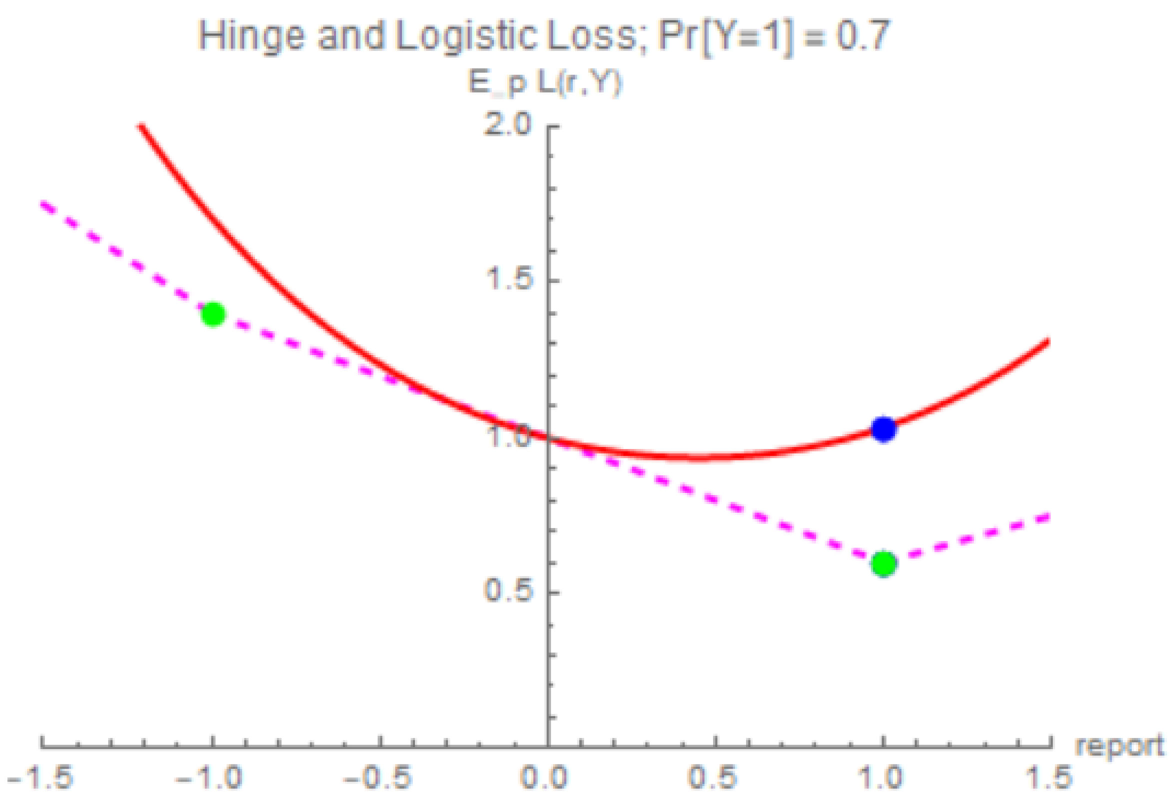
Want a way to get a calibrated surrogate from a given discrete loss.

We say $L : \mathbb{R}^d \to \mathbb{R}^n_+$ embeds $\ell : \mathcal{R} \to \mathbb{R}^n_+$ if there exists an injective embedding $\varphi : \mathcal{R} \to \mathbb{R}^d$ such that

i. For all $r \in \mathcal{R}$, we have $L(\varphi(r)) = \ell(r)$.

ii. For all $p \in \Delta_{\mathcal{Y}}$,

$$r \in \operatorname{argmin}_{r' \in \mathcal{R}} \langle \ell(r'), p \rangle \\ \Leftrightarrow \\ \varphi(r) \in \operatorname{arginf}_{u' \in \mathbb{R}^d} \langle L(u'), p \rangle$$



Example: Hinge loss (pink, dashed) embeds twice discrete 0-1 loss, (green dots) but logistic loss does not.

$\mathcal{R} = \{-1, 1\}$

Results

Prop 1: L embeds ℓ if and only if $\underline{L} = \underline{\ell}$.

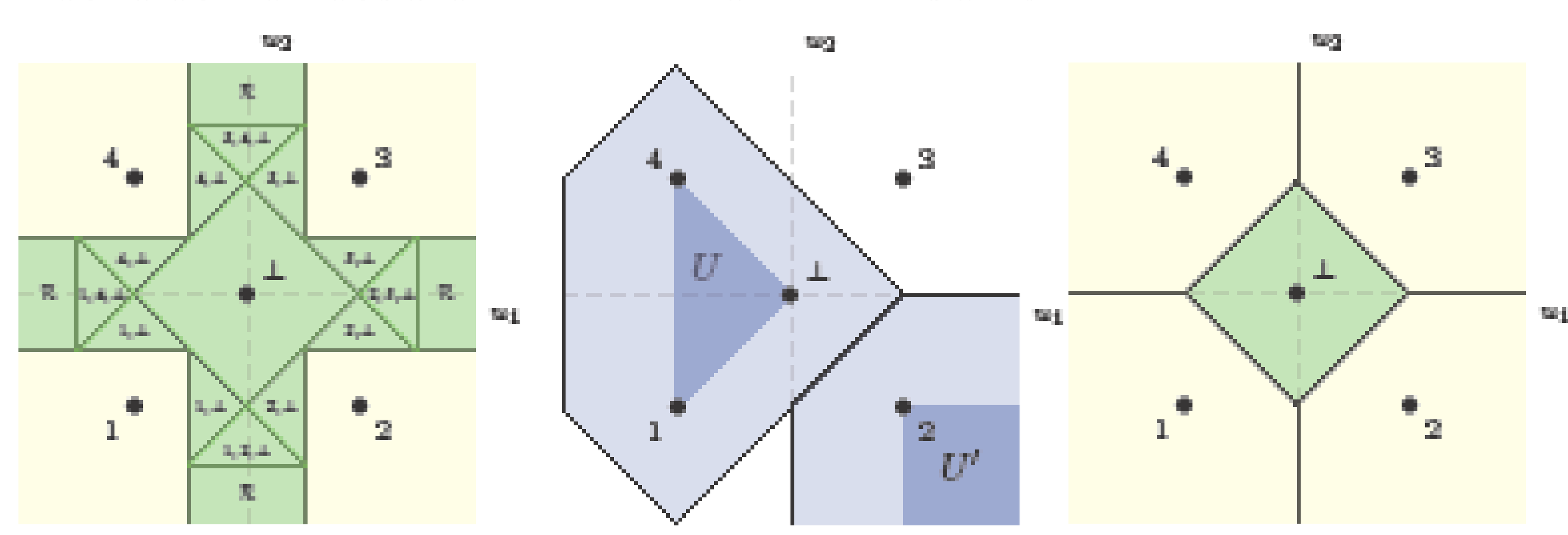
Prop 2: A loss L embeds a discrete loss if and only if $\underline{L}$ is polyhedral.

Theorem 1: Every polyhedral loss L embeds a discrete loss.

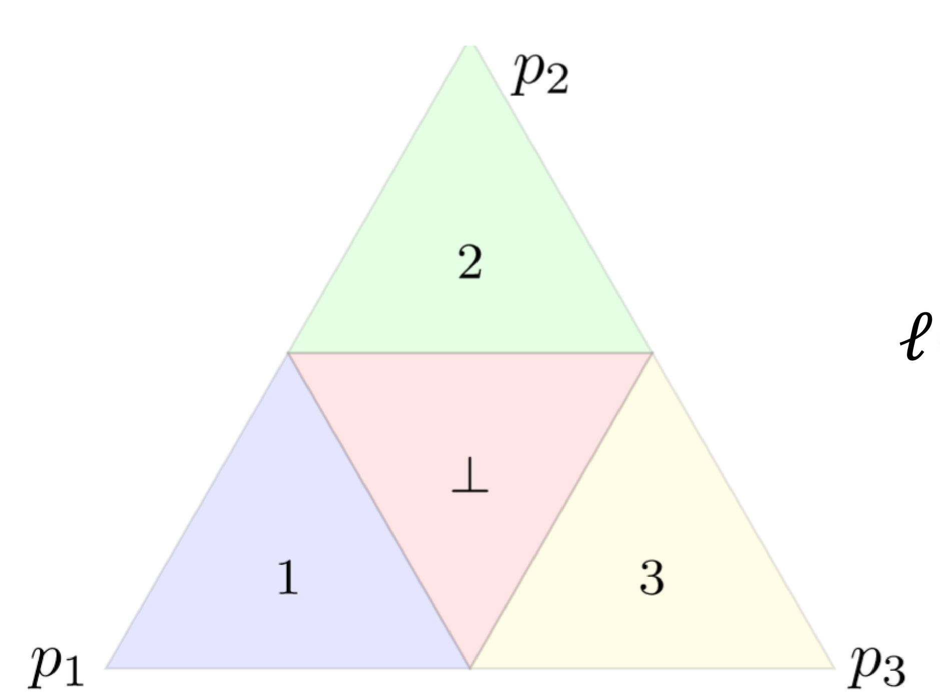
Theorem 2: Every discrete loss ℓ is embedded by a polyhedral loss.

Calibrated Links

Theorem 3: L embeds ℓ implies there is a calibrated link from L to ℓ.



Example: Abstain Loss



$$\ell(r, y) = \begin{cases} 0 & r = y \\ \frac{1}{2} & r = \bot \\ 1 & r \neq y, r \neq \bot \end{cases}$$

Future Work

<u>Embedding dimension</u>: Can we minimize the dimension of the input to the calibrated surrogate?