An Embedding Framework for Calibrated Polyhedral Surrogates

Jessie Finocchiaro, Rafael Frongillo, Bo Waggoner

University of Colorado Boulder

Overview

In classification-like problems, we seek easy-to-optimize surrogate losses. Current surrogates are typically designed in an ad hoc manner, with consistency not guaranteed. We introduce an embedding framework to design convex surrogates for a given discrete loss.

Setting

$\mathcal{R}, \mathcal{Y}$	Finite prediction/outcome sets
$\ell: \mathcal{R} \to \mathbb{R}^n_+$	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
$\psi: \mathbb{R}^d \to \mathcal{R}$	Link function

Let original loss ℓ, proposed surrogate L, and link function ψ be given. We say (L, ψ) is **calibrated** with respect to ℓ if, for all $p \in \Delta_\mathcal{Y}$,

$$\inf_{u \in \mathbb{R}^d : \psi(u) \notin arg\,min_u \langle p, \ell(u) \rangle} \langle p, L(u) \rangle > \inf_{u \in \mathbb{R}^d} \langle p, L(u) \rangle .$$

Embedding Framework

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

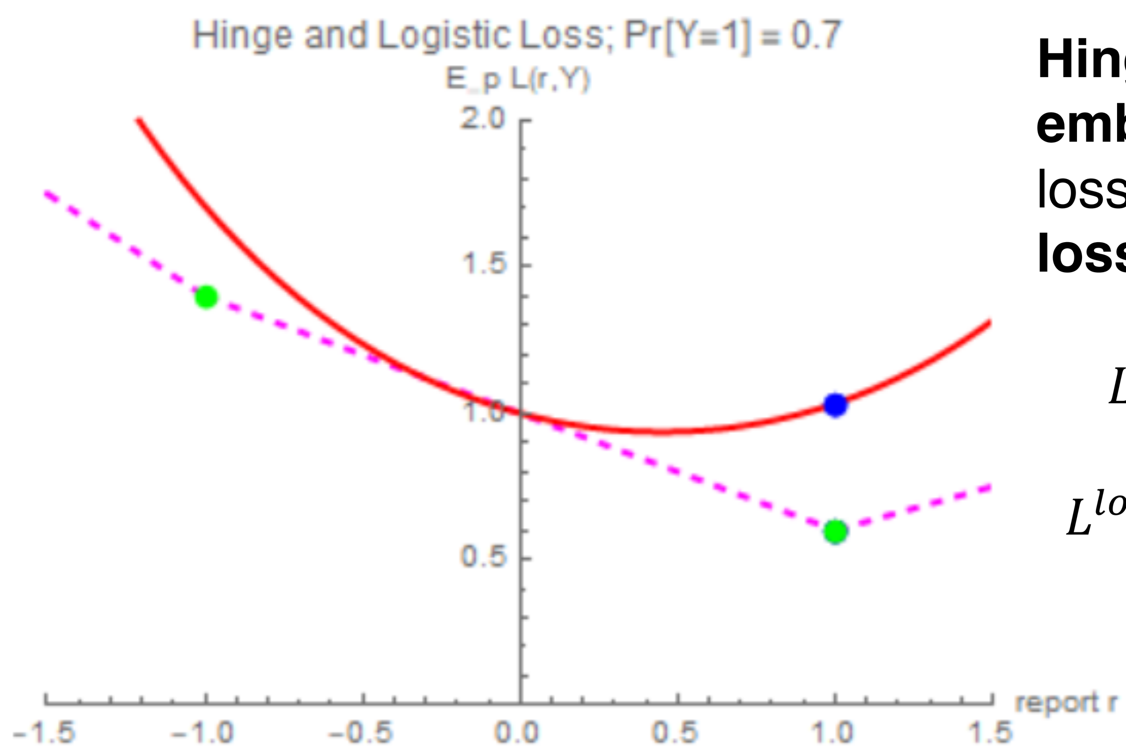
i. For all $r \in \mathcal{R}$, we have $L(\varphi(r)) = \ell(r)$.

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

$$r \in \arg\min_{r' \in \mathcal{R}} \langle \ell(r'), p \rangle \Leftrightarrow \varphi(r) \in \arg \inf_{u' \in \mathbb{R}^d} \langle L(u'), p \rangle$$

The prediction dimension *d* above refers to the **embedding dimension**, which we would like to minimize.

Example 1: Hinge Loss



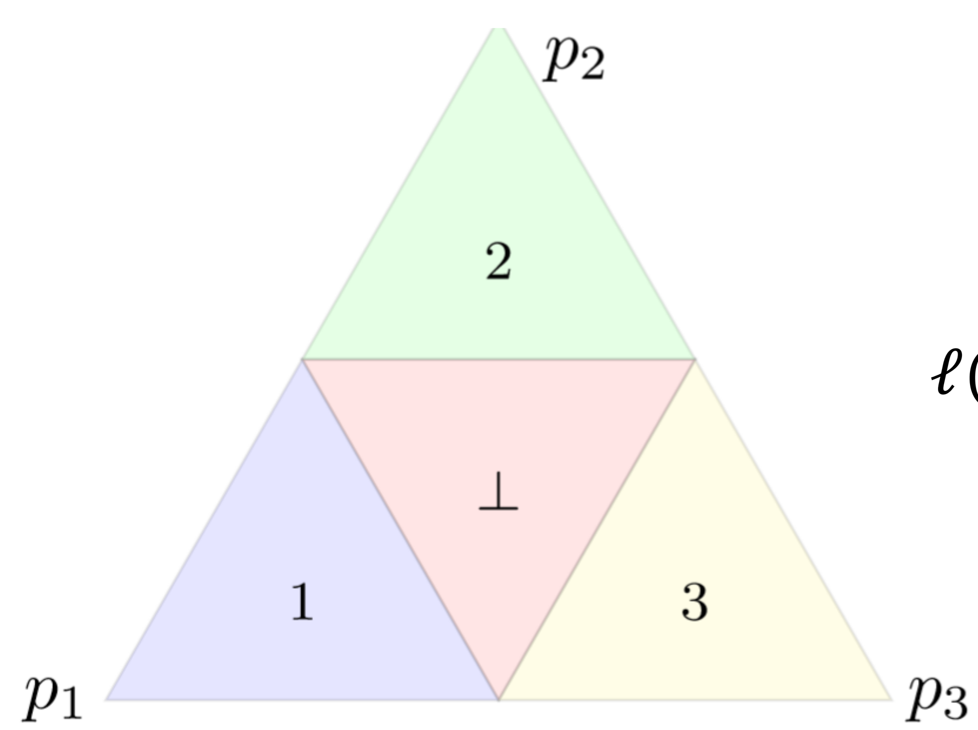
Hinge loss (pink, dashed) **embeds** twice discrete 0-1 loss, (green dots) but **logistic loss** (red, solid) **does not**.

$$L^{hinge}(r, y) = (1 - ry)_+$$

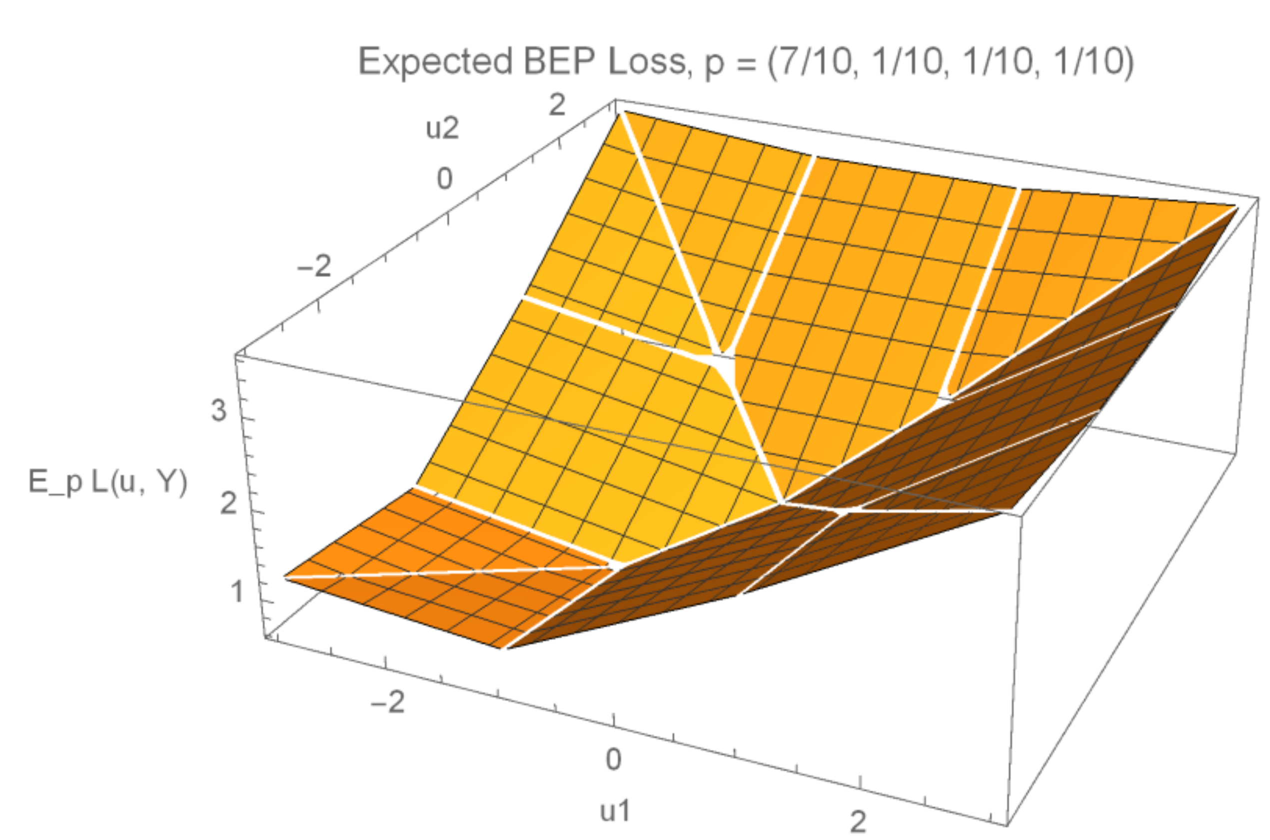
$$L^{logistic}(r, y) = \lg(1 + e^{-ry})$$

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

Example 2: Abstain Loss



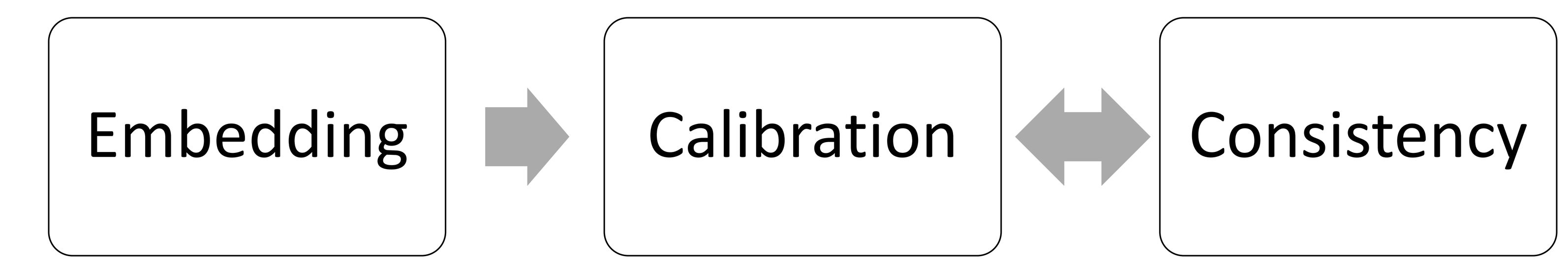
$$\ell(r, y) = \begin{cases} 0 & r = y \\ \frac{1}{2} & r = \bot \\ 1 & r \neq y, r \neq \bot \end{cases}$$



BEP Surrogate $L(u, y) = (\max_{j \in [d]} B_j(y) u_j + 1)_+$ is calibrated for abstain loss.

(Ramaswamy, Tewari, Agarwal. (2018.) Consistent algorithms for multiclass classification with an abstain option. In *Electronic Journal of Statistics*)

Context of results



This work [FFW19, FFW20]

Previous work [BJM06, TB07, RTA16]

Results

Theorem 1 [FFW19]: Every polyhedral loss L embeds a discrete loss.

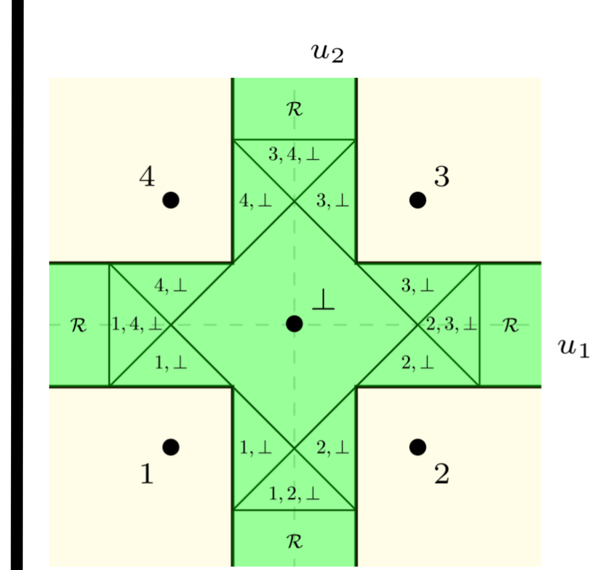
Theorem 2 [FFW19]: Every discrete loss ℓ is embedded by a polyhedral loss.

Theorem 12 [FFW20]: For d = 1, calibration and embedding are equivalent.

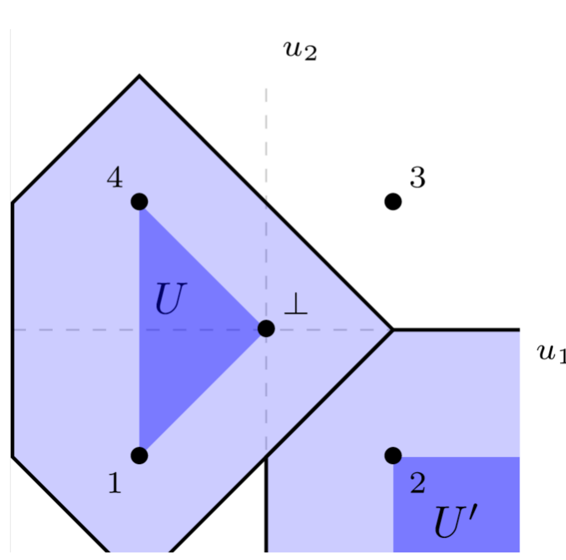
Theorem 18 [FFW20]: A surrogate is d-embeddable if and only if quadratic feasibility program has solution.

Calibrated Links

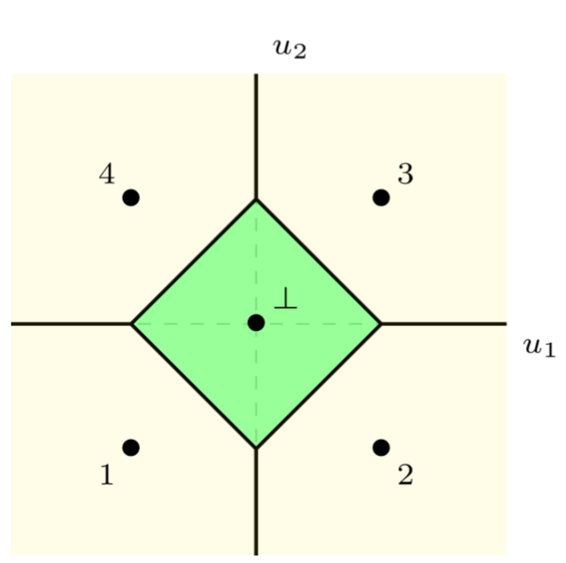
Theorem 3 [FFW19]: L embeds ℓ implies there is a calibrated link from L to ℓ.



Possible calibrated link values by constructing link with $||\cdot||_\infty$ and $\epsilon = 1/2$.



Examples of U sets that are used to calculate the calibrated link for the BEP embedding.



Calibrated link using $||\cdot||_1$ and $\epsilon = 1$.

Questions? @jessie_fin or email jefi8453@colorado.edu