# An Embedding Framework for Calibrated Polyhedral Surrogates

Jessie Finocchiaro[1], Rafael Frongillo[1], Bo Waggoner[1,2]

[1]University of Colorado Boulder, [2]Microsoft Research

## Overview

In classification-like problems, we often seek surrogate losses to escape the intractability of directly optimizing the given discrete loss (e.g. 0-1 loss for classification). Current surrogates are typically designed in an ad hoc manner, with calibration not guaranteed. Motivated by recent **polyhedral (piecewise linear convex)** surrogate losses in the literature, we introduce an embedding framework to design calibrated, polyhedral 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 |

## Properties and Calibration

**Definition 1:** A **property** $\Gamma: \Delta_\mathcal{Y} \rightrightarrows \mathcal{R}$ is a function mapping probability distributions to sets of reports.

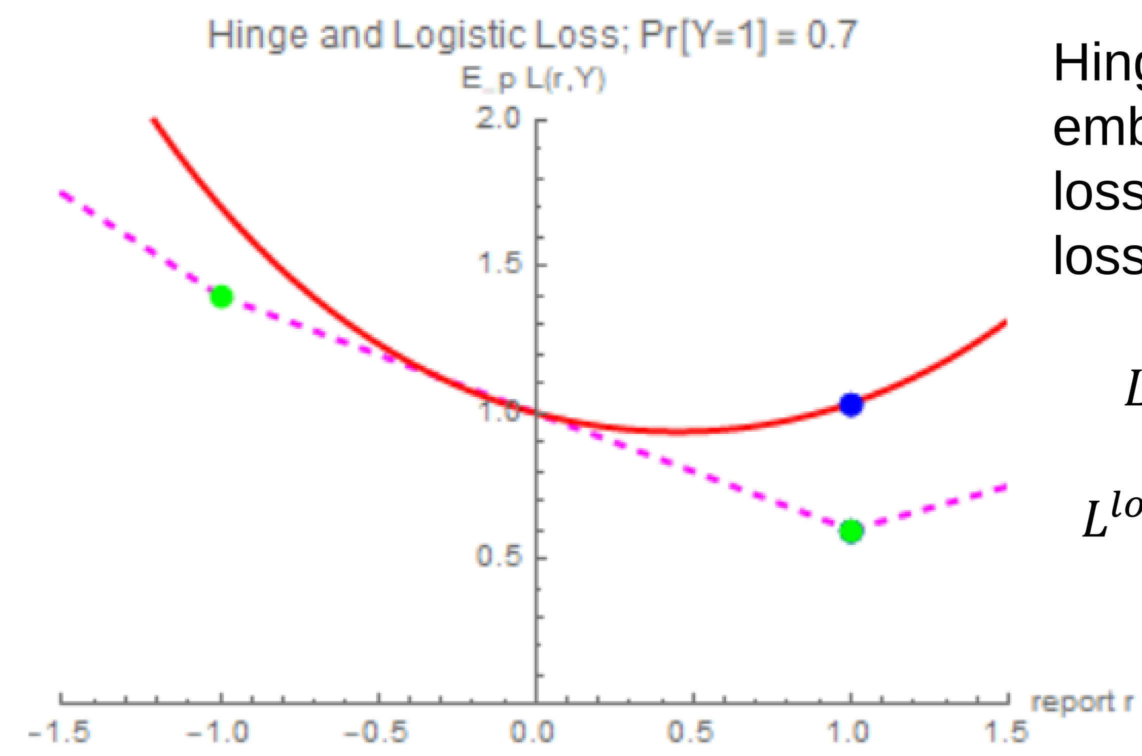
**Definition 2:** We say a loss $L$ **elicits** the property $\Gamma$ if, for all $p \in \Delta_\mathcal{Y}$, we have

$$\Gamma(p) = \arg\min_{r \in \mathcal{R}} \langle p, L(r) \rangle.$$

**Definition 3:** Let original loss $\ell$ eliciting $\Gamma$, proposed surrogate $L$, and link function $\psi$ be given. We say $(L, \psi)$ is **calibrated** with respect to $\ell$ if, for all $p \in \Delta_\mathcal{Y}$,

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

## Example 1: Hinge Loss



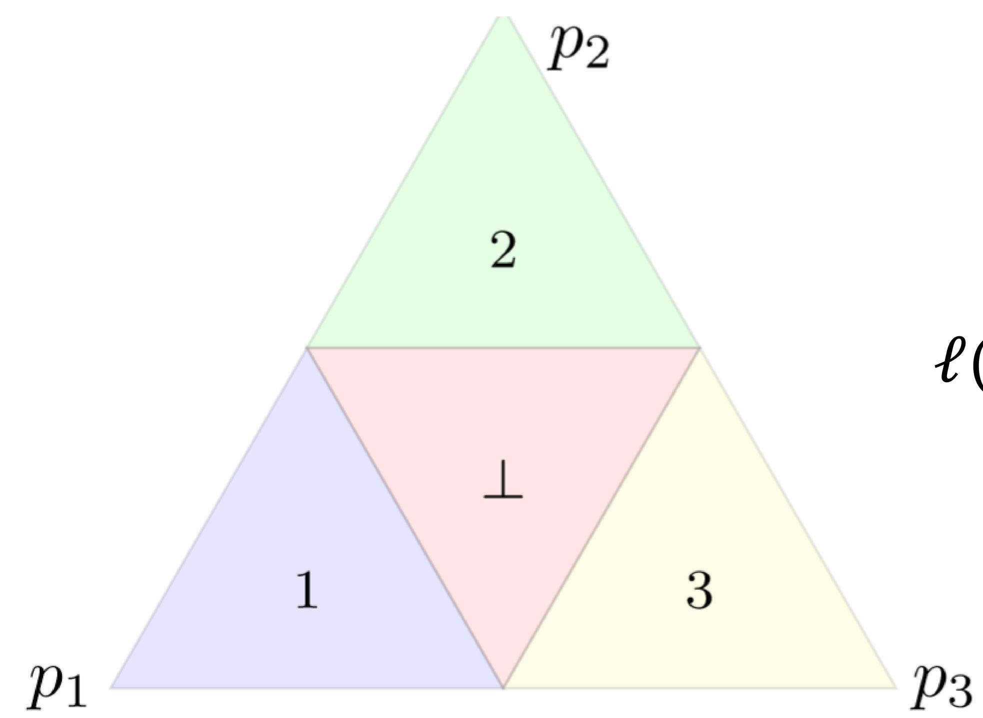

Hinge loss (pink, dashed) embeds twice discrete 0-1 loss, (green dots) but logistic loss (red, solid) does not.

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

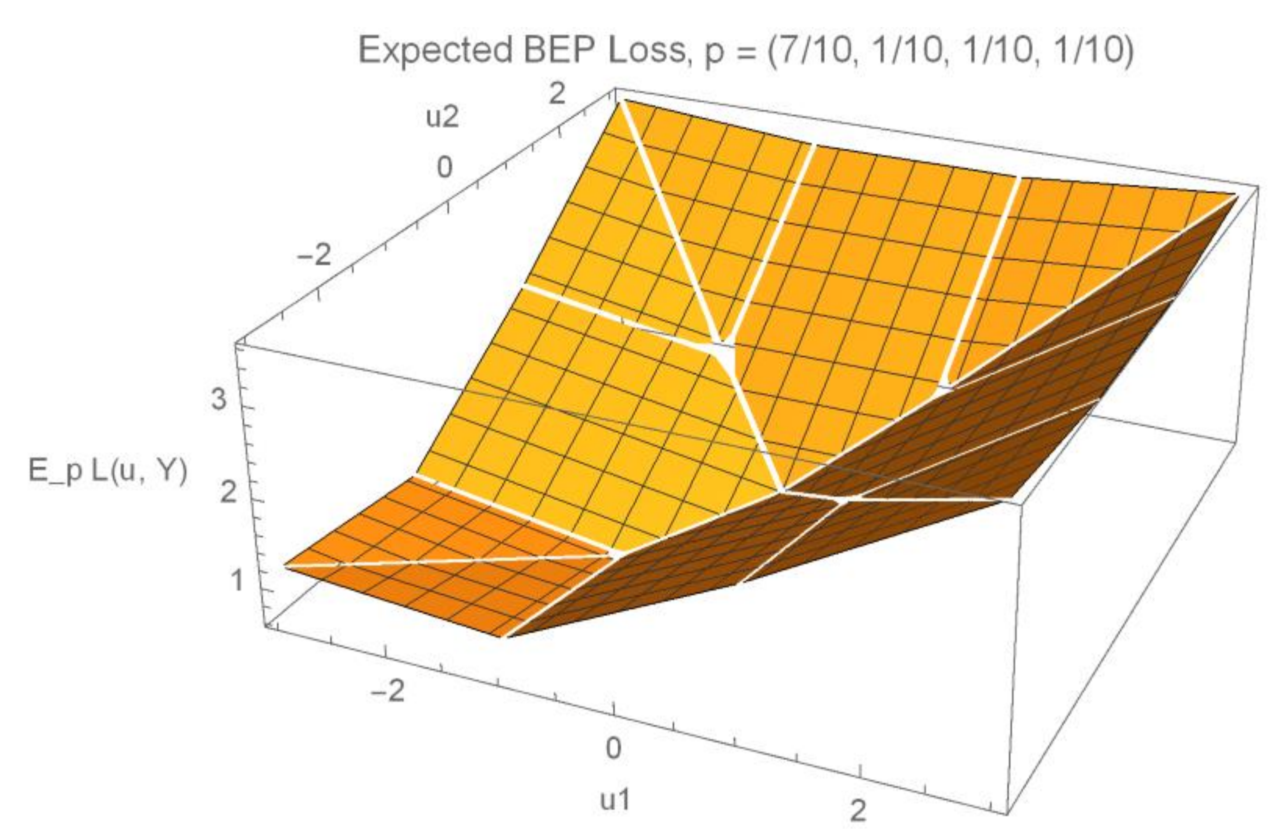
$$L^{logistic}(r, y) = \lg(1 + e^{-ry})$$

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

## Example 2: Abstain Loss




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




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

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

## 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 $p \in \Delta_\mathcal{Y}$,

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

## Results

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

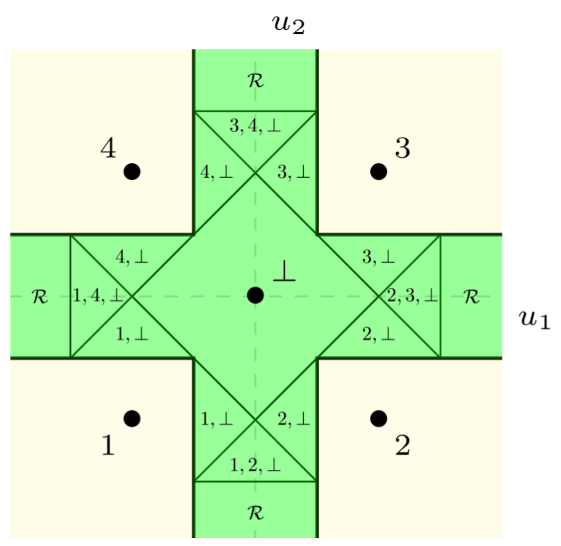
**Prop 2:** A loss $L$ embeds a discrete loss if and only if $\underline{L}$ is polyhedral.

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

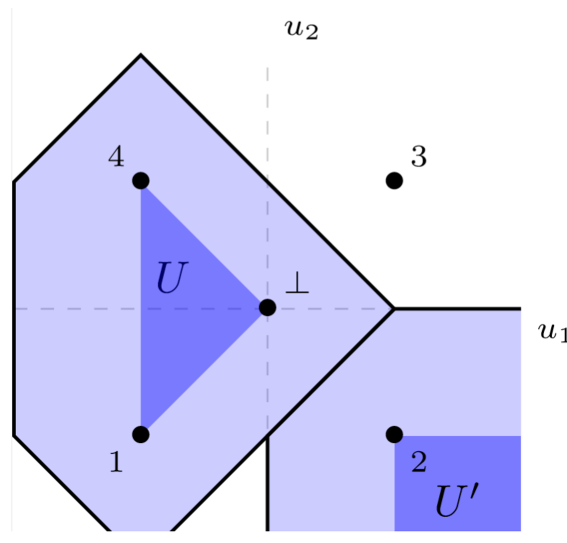
**Theorem 2:** Every discrete loss $\ell$ is embedded by a polyhedral loss.

## Calibrated Links

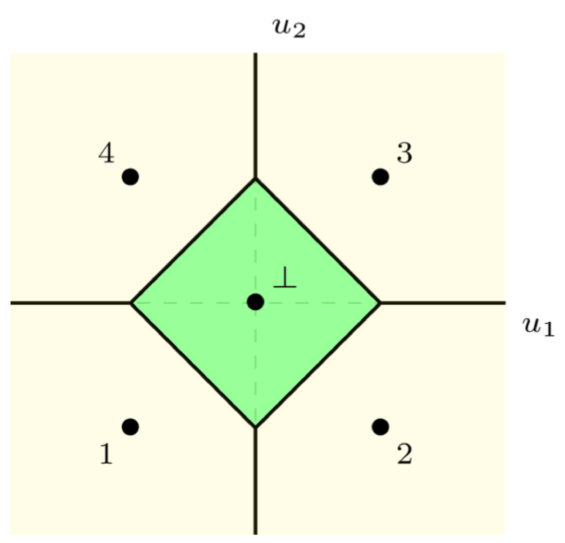
**Theorem 3:** $L$ embeds $\ell$ implies there is a calibrated link from $L$ to $\ell$.




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 $\epsilon = 1$.

## Future Work

**Embedding dimension**: Can we find the minimum dimension $d$ such that $\mathrm{L}: \mathbb{R}^d \to \mathbb{R}^n_+$ is calibrated with respect to a given $\ell$?