Unifying Lower Bounds on Prediction Dimension via Convex Surrogates

Jessie Finocchiaro, Rafael Frongillo, Bo Waggoner
University of Colorado Boulder

Introduction

A loss function is called a *surrogate* when it is used to solve a related, but not identical "target" problem of interest. Selecting a hypothesis by minimizing surrogate risk is one of the most widespread techniques in supervised ML. There are two main reasons why a surrogate loss is necessary: (I) the target problem is to minimize a loss, which does not satisfy some desiderata such as continuity or convexity, or (II) the target problem is some statistic, and no loss is given. We present a framework that applies to both settings simultaneously and apply it to study *low-dimensional* convex surrogate losses.

Notation

$\mathcal{R}, \mathbb{R}^d$	Reports: discrete, surrogate
$\mathcal{Y}, p \sim \mathcal{P} \subseteq \Delta_{\mathcal{Y}}$	Outcomes, set of distributions
$\ell(r, y), L(u, y)$	Discrete, surrogate loss
$\Gamma : \mathcal{P} \to \mathbb{R}^d$	Surrogate property
$\forall p \in \mathcal{P}, \Gamma(p) = argmin_{u \in \mathbb{R}^d} \mathbb{E}_{Y \sim p} L(u, Y)$	Loss elicits property
$\Gamma_r = \{ p \in \mathcal{P} \mid r \in \Gamma(p)\}$	Level set of property

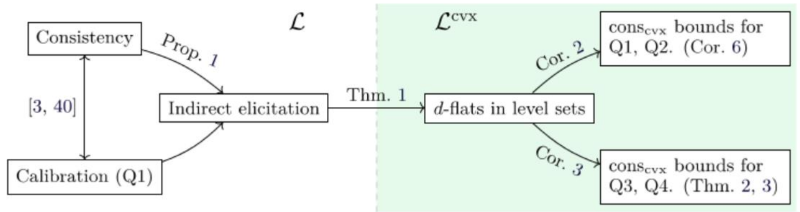
Four types of problem

	Target loss	*Target statistic*
Discrete prediction	**Q1**, e.g. classification	**Q2**, e.g. hierarchical classification
Continuous estimation	**Q3**, e.g. least-squares regression	**Q4**, e.g. variance estimation

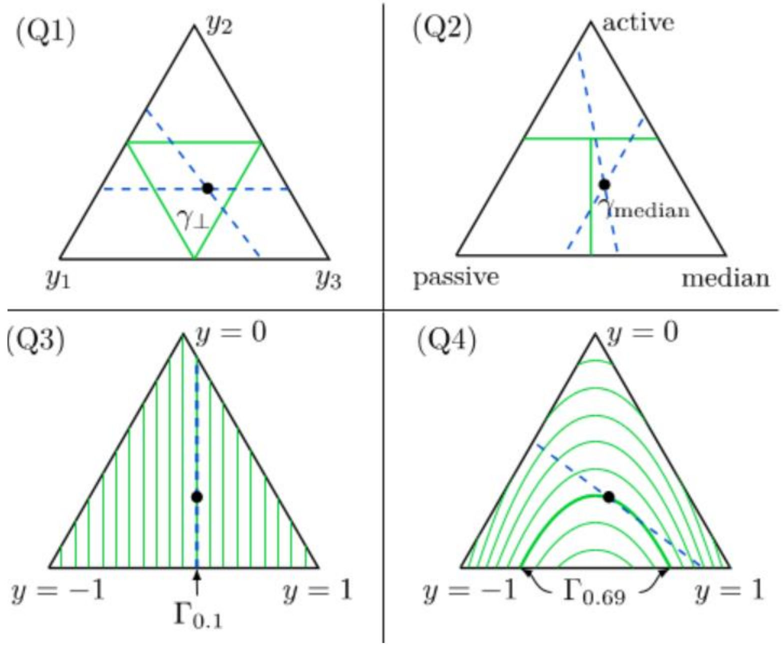
Goal

Given a target loss or statistic, lower bound on the dimension d of any consistent convex surrogate $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}$.

Roadmap of Results



Example Application of Results



Definitions

Convex prediction dimension: The minimum dimension d such that a convex $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}_+$ is consistent/calibrated/elicits the task at hand.

The ***d*-flat** is a nonempty set $F = \{p \in \mathcal{P} \mid \mathbb{E}_p V = \vec{0}\}$ parameterized by the measurable function $V : \mathcal{Y} \to \mathbb{R}^d$.

Lower bound via *d*-flats

If a surrogate elicits a property Γ, then for all $p \in \mathcal{P}, u \in \Gamma(p)$, there is a d-flat F containing p such that $F \subseteq \{p \in \mathcal{P} \mid u \in \Gamma(p)\}$

New lower bounds

Continuous estimation problems (Q3,Q4):
Let dim(Γ) = convex consistency dimension.

- dim(Conditional Value at Risk) ≥ 2
  - Solves an open problem in statistics/finance
- dim(Mode) = dim(Modal interval) = ∞
  - Solves an open problem in statistics

Discrete Prediction (Q1, Q2):
Subsume [RTA16] feasible subspace dimension result, and in some cases strict improvement.

Unifying Lower Bounds on Prediction Dimension via Convex Surrogates

Jessie Finocchiaro, Rafael Frongillo, Bo Waggoner
University of Colorado Boulder