Consistent Polyhedral Surrogates for Top-k Classification and Variants

Jessie Finocchiaro, Rafael Frongillo, Emma Goodwill, Anish Thilagar

July 19th, 2022

ICML

Background: top-k loss

Use cases: information retrieval, image classification, etc.

$S \in \{1, \ldots, n\}$, $|S| = k$

$$\ell_k(S, y) = \begin{cases} 0 & y \in S \\ 1 & y \notin S \end{cases} .$$

Discrete: optimizing is hard $\rightarrow$ use a continuous surrogate.

Example: Surrogates for 0-1 loss: hinge loss, log loss

Desirable surrogates: convex, consistent

Previous top-k surrogates

- Cross-entropy: smooth, learns entire distribution

  smooth surrogates tend to have worse regret bounds (Frongillo and Waggoner 2021)

- Polyhedral surrogates: inconsistent (so far...)

  Lapin, Hein, and Schiele 2016, Yang and Koyejo 2020

Questions

1. What do those inconsistent polyhedral surrogates "do"?
2. Does a consistent polyhedral top-k surrogate exist?

   previous literature suggests: no

Results via Embedding Framework

Questions

1. What do those inconsistent polyhedral surrogates "do"?
2. Does a consistent polyhedral top-k surrogate exist?

   previous literature suggests: no

Answers

1. Consistent for certain variants of the top-k problem.

   more relevant than top-k in certain settings
2. Yes!

$$L_k(u, y) = \max_{m \in 1,\ldots,n} \left\{ \frac{\sigma_m(u)}{m} + \left(1 - \frac{k}{m}\right)_+ \right\} - u_y \ .$$

Both results rely on the embedding framework derived in Finocchiaro, Frongillo, and Waggoner 2022.