Ranking: Consistency and Rates

Dany Haddad

July 2020

1 Introduction

This is a work in progress document discussing the consistency of common convex surrogate functions for ranking as well as the convergence rate of the corresponding empirical risk. First we summarize the main points of Duchi et al. in this regard [**?**, **?**]. Then we discuss some relevant research directions, in particular, we are interested in achieving fast convergence rates for ranking applications. [1]

2 Literature Review

2.1 The Asymptotics of Ranking Algorithms

2.1.1 Overview

The authors first introduce the ranking problem and describe a pairwise loss function, equivalent to the pairwise 0-1 ranking loss. They then define several forms of consistency for ranking problems and unify them under some conditions (theorem 1). Surprisingly, common surrogates for the pairwise 0-1 ranking loss function are shown to not be consistent.

The next section discusses how aggregating the pairwise preferences into lists allows us to define consistent and tractable surrogate loss functions. The key idea is that the surrogates must be order preserving. [2]

Given that our new surrogate loss functions consist of an aggregation step, it is natural to consider the aggregation in batches. In particular, the authors describe the loss function as a sum of U-statistics. However, this is seems mostly to simplify the analysis rather than to leverage the fact that U-statistics are UMVUEs. [3] The authors then prove a ULLN for the risk functional they define and show that it is consistent for the underlying listwise losses.

[1] I also point out confusion and questions I have in the footnotes.

[2] The authors describe inconsistency as arising due to a lack of complete preference information. This point is not entirely clear to me since the proof of inconsistency seems to depend on the variability in preference information rather than the lack of information. See [**?**] for a discussion on how a lack of normalization leads to "non-robustness".

[3] The authors mention that they introduce the U-statistics based risk functional to be able to analyze the setup without overly detailed knowledge of the surrogate loss. But they already

I still have not spent sufficient time on the experimental section to provide an overview.

2.2 Proving a ULLN for the U-statistics Based Risk Functional

assume the surrogate is lipschitz continuous and bounded. What more information would we need to conduct the analysis without taking the U-statistics route?