Surrogate Functions for Maximizing Precision at the Top

Purushottam Kar*, Harikrishna Narasimhan*, and Prateek Jain

*Microsoft Research, Bengaluru, INDIA
†Indian Institute of Science, Bengaluru, INDIA

The Goal

Scalable routines for provable maximization of precision at the top of ranked lists

Applications

Drug Discovery
Recommendation Systems
Document Tagging

What is a good Surrogate for prec@k?

- Convexity (CV)
  \( \ell_{\text{prec@k}}(s) \) is convex over \( \mathbb{R}^n \)
- Upper-bounding Property (UB)
  \( \ell_{\text{prec@k}}(s) \geq \text{prec@k}(s) \) \( \forall s \)
- Tight under a Margin (TuM)
  For classes of score vectors \( s \) satisfying an appropriate margin condition
  \( \min_{s \in S} \ell_{\text{prec@k}}(s) = \min_{s \in S} \text{prec@k}(s) \)

A Notion of Margin for prec@k

Classification Margin

(Weak) prec@k Margin

Surrogates for prec@k

Ramp Surrogate
\[
\text{ramp}_{\text{prec@k}}(s) = \max_{|y| \leq k} \left\{ \frac{\text{prec@k}(y)}{\text{prec@k}(0)} + \sum_{i \in k} y_i s_i \right\} - \max_{|y| \leq k} \frac{\sum_{i \in k} y_i s_i}{\text{prec@k}(0)}
\]

Avg Surrogate
\[
\text{avg}_{\text{prec@k}}(s) = \frac{1}{|y|} \left( \text{prec@k}(0) + \frac{\sum_{i \in k} y_i s_i}{\text{prec@k}(0)} \right) \frac{1}{|y|} \sum_{i \in k} (1 - y_i) s_i
\]

Penalize score vectors that don't give k relevant objects the highest scores

Relax the Ramp surrogate or else add corrections to the struct-SVM surrogate

Lemma: The avg (ramp) surrogate is tight for any class of score vectors \( s \) that contains a score vector realizing a unit (weak) prec@k margin.

Question I

Convex, Upper-bounding and Conditionally Consistent surrogates for prec@k

Question II

Scalable optimization of prec@k in large-scale and streaming data settings

Methodology

- Given n objects \( (x_i, y_i), y_i \in \{0, 1\} \)
- Assign scores \( s = (s_1, s_2, \ldots, s_n) \)
- Predict top-k scoring objects as relevant
- Learn models that predict good score vectors
- Learning on streaming data?

Experiments

- Gradient descent-based approach GD@k based on surrogates
- Mini-batch versions of PERCEPTRON@k and GD@k
- Mistake/generalization bounds via OTB/UC

Full Paper: http://tinyurl.com/p3vjg7