Ranking Data with Continuous Labels through Oriented Recursive Partitions

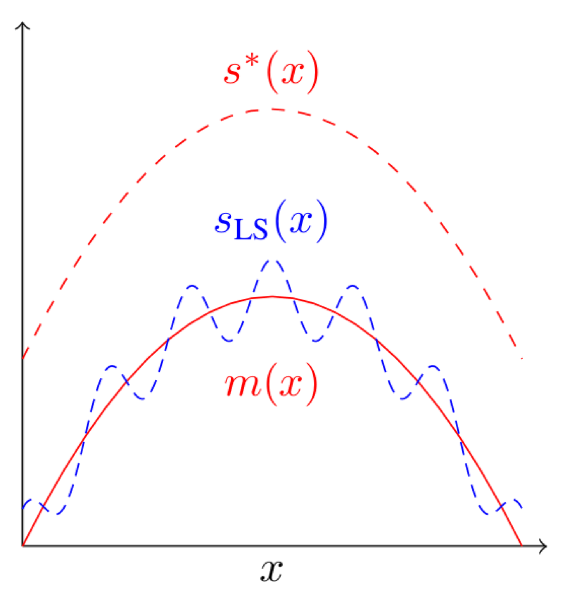
Stephan Clémençon Mastane Achab
LTCI, Télécom ParisTech

Baseline: bipartite ranking

- n observations: $(X_i, Y_i) \in \mathcal{X} \times \{-1, 1\}$, where $\mathcal{X} \subset \mathbb{R}^d$
- Goal: build a scoring function $s : \mathcal{X} \to \mathbb{R}$ s.t. $s(X)$ and Y $\nearrow$ or $\searrow$ together with large probability
- Functional criterion: maximize $\forall \alpha \in [0,1]$, $ROC_s(\alpha) = 1 - G_s \circ (1 - H_s^{-1})(1-\alpha)$, with G_s and H_s cdf of $(s(X)|Y = +1)$ and $(s(X)|Y = -1)$
- More convenient to maximize a scalar summary criterion: $AUC(s) = \int_{\alpha=0}^{1} ROC_s(\alpha) d\alpha = \mathbb{P}(s(X) < s(X')|Y < Y') + \frac{1}{2}\mathbb{P}(s(X) = s(X')|Y < Y')$
- Optimal scores: $s^*(x) = \mathbb{P}(Y = 1|X = x)$ and stricly increasing transforms
- TreeRank: CART-like algorithm maximizing empirical AUC $\to$ produces piecewise constant scoring function

Our problem: continuous ranking

- Continuous label $Y \in [0,1]$
- Goal: build a scoring function s "good" for any bipartite subproblem at level $y \in [0,1]$, with $Z_y = -1$ if $Y < y$ and $Z_y = +1$ if $Y > y$
- Continuum of functional sub-criteria: $\forall y \in [0,1]$, maximize $\forall \alpha \in [0,1]$, $ROC_{s,y}(\alpha)$
- Aggregated criteria: $IROC_s(\alpha) = \int_{y=0}^{1} ROC_{s,y}(\alpha) F_Y(dy)$ and $IAUC(s) = \int_{\alpha=0}^{1} IROC_s(\alpha) d\alpha = \mathbb{P}(s(X) < s(X')|Y < Y'' < Y') + \frac{1}{2}\mathbb{P}(s(X) = s(X')|Y < Y'' < Y')$
- Existence of optimal scores: regression model $Y = h(X) + \epsilon$, exponential families $\to$ optimal scores: $m(x) = \mathbb{E}[Y|X = x]$ and stricly increasing transforms



- regression function $m(x)$ is optimal
- s_{LS} good for least squares regression but not for ranking
- $s^*(x)$ is also optimal even if large MSE

References

- S. Robbiano S. Clémençon and N. Vayatis. Ranking data with ordinal labels: optimality and pairwise aggregation. Machine Learning, 91(1):67–104, 2013.
- S. Clémencçon and N. Vayatis. Tree-based ranking methods. IEEE Transactions on Information Theory, 55(9):4316–4336, 2009.
- L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth and Brooks, 1984.

CRank algorithm

- Same idea as TreeRank
- In CRank, IAUC plays the same role as AUC in TreeRank

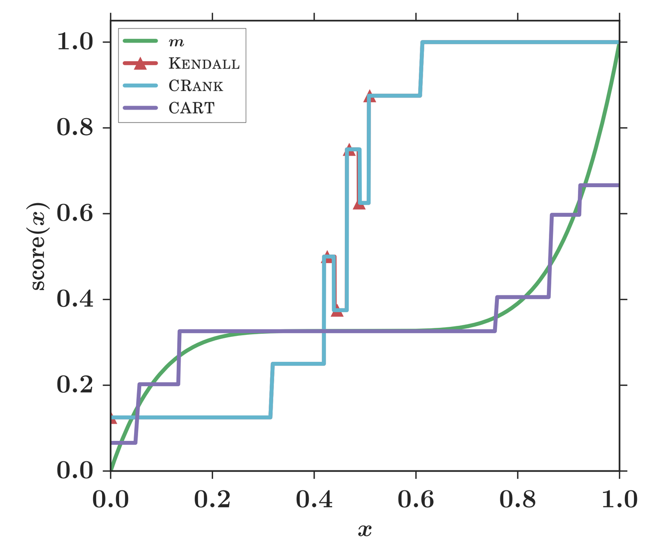
CRank

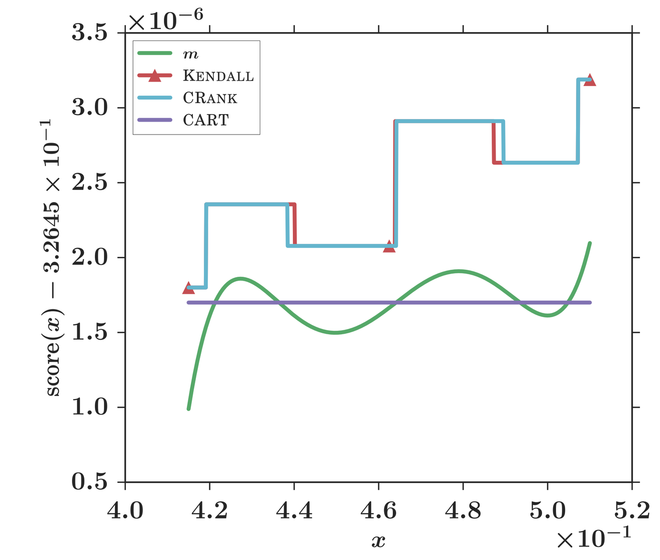
1: **Input.** Training data $\{(X_i, Y_i)\}_{i=1}^{n}$, tree depth $D \geq 1$.
2: **Initialization.** Set $\mathcal{C}_{0,0} = \mathcal{X}$.
3: **Iterations.** For $d = 0, \ldots, D-1$ and $k = 0, \ldots, 2^d - 1$,
   (a) Find the best sub-rectangle $C_{d+1,2k}$ of rectangle $C_{d,k}$ in the empirical IAUC sense.
   (b) Then, set $C_{d+1,2k+1} = C_{d,k} \setminus C_{d+1,2k}$.
4: **Output.** After D iterations, we get the piecewise constant scoring function:

$$s_D(x) = \sum_{k=0}^{2^D-1} (2^D - k)\mathbb{I}\{x \in C_{D,k}\}.$$

Numerical experiments

- Regression model without noise: $Y = m(X)$, with X and Y valued in $[0,1]$ and m a polynomial function.
- Critical window $I = [0.4, 0.5]$ where m slightly oscillates and $\mathbb{P}(Y \in I) = 0.8$.
- Training on $\{(X_i, Y_i)\}_{i=1}^{n_{\text{train}}}$ with $Y_i = m(X_i)$ and $n_{\text{train}} = 100$ with tree depth $D = 3$ $\to$ then test on $n_{\text{test}} = 2000$ new iid copies of X.





	IAUC	Kendall τ	MSE
CRank	0.95	0.92	0.10
Kendall	0.94	0.93	0.10
CART	0.61	0.58	7.4×10^{-4}

Table 1: IAUC, Kendall τ and MSE empirical measures on test set