Ranking Median Regression: Learning to Order through Local Consensus

DISCML 2017

Stephan Clémençon, **Anna Korba**, Eric Sibony

December 8, 2017

Télécom ParisTech

Introduction

- Let $\{1, \ldots, n\}$ a set of items to be ranked.
- A full ranking $i_1 \succ \cdots \succ i_n$ is seen as a permutation $\sigma \in \mathfrak{S}_n$ that maps an item i_j to its rank j.

Ex: $2 \succ 3 \succ 1 \Leftrightarrow \sigma(2) = 1, \sigma(3) = 2, \sigma(1) = 3$.

Introduction

- Let $\{1, \dots, n\}$ a set of items to be ranked.
- A full ranking $i_1 \succ \dots \succ i_n$ is seen as a permutation $\sigma \in \mathfrak{S}_n$ that maps an item i_j to its rank j.

Ex: $2 \succ 3 \succ 1 \Leftrightarrow \sigma(2) = 1, \sigma(3) = 2, \sigma(1) = 3$.

Suppose we observe $((X_1, \Sigma_1), \dots, (X_N, \Sigma_N)$ i.i.d. copies of the pair $(X,\ \Sigma)$, where

- $X \sim \mu$, where μ is a distribution on some feature space $\mathcal{X}$
- $\Sigma \sim P_X$, conditional prob. distr. on the symmetric group $\mathfrak{S}_n$

Ex: Users i with characteristics X_i order items by order of preference Σ_i.

Introduction

- Let $\{1, \ldots, n\}$ a set of items to be ranked.
- A full ranking $i_1 \succ \cdots \succ i_n$ is seen as a permutation $\sigma \in \mathfrak{S}_n$ that maps an item i_j to its rank j.

Ex: $2 \succ 3 \succ 1 \Leftrightarrow \sigma(2) = 1, \sigma(3) = 2, \sigma(1) = 3$.

Suppose we observe $((X_1, \Sigma_1), \ldots, (X_N, \Sigma_N)$ i.i.d. copies of the pair $(X,\ \Sigma)$, where

- $X \sim \mu$, where μ is a distribution on some feature space $\mathcal{X}$
- $\Sigma \sim P_X$, conditional prob. distr. on the symmetric group $\mathfrak{S}_n$

Ex: Users i with characteristics X_i order items by order of preference Σ_i.

Goal of Ranking Regression

Learn a predictive ranking rule $s : \mathcal{X} \to \mathfrak{S}_n$ which given a random vector X (e.g characteristics of an user), predicts the order/permutation Σ on the n items (e.g its true preferences).

Ranking Median Regression

Formally, find $s \in \mathcal{S}$ minimizing the following (theoretical) **risk**:

$$\mathcal{R}(s) = \mathbb{E}_{X \sim \mu, \Sigma \sim P_X} \left[d_\tau \left(s(X), \Sigma \right) \right], \tag{1}$$

where d is the Kendall's tau distance, i.e for all $(\sigma, \sigma') \in \mathfrak{S}_n^2$:

$$d_\tau(\sigma, \sigma') = \sum_{i<j} \mathbb{I}\{(\sigma(j) - \sigma(i)) \cdot (\sigma'(j) - \sigma'(i)) < 0\}$$

Ranking Median Regression

Formally, find $s \in \mathcal{S}$ minimizing the following (theoretical) **risk**:

$$\mathcal{R}(s) = \mathbb{E}_{X \sim \mu, \Sigma \sim P_X}\left[d_\tau\left(s(X), \Sigma\right)\right], \tag{1}$$

where d is the Kendall's tau distance, i.e for all $(\sigma, \sigma') \in \mathfrak{S}_n^2$:

$$d_\tau(\sigma, \sigma') = \sum_{i<j} \mathbb{I}\{(\sigma(j) - \sigma(i)) \cdot (\sigma'(j) - \sigma'(i)) < 0\}$$

Definition

For any $\sigma \in \mathfrak{S}_n$, the risk of σ is $L(\sigma) = \mathbb{E}_{\Sigma \sim P}[d(\Sigma, \sigma)]$. A **Kemeny median** is any

$$\sigma_P^* = \operatorname*{argmin}_{\sigma \in \mathfrak{S}_n} L(\sigma) \tag{2}$$

Ranking Median Regression

Formally, find $s \in \mathcal{S}$ minimizing the following (theoretical) **risk**:

$$\mathcal{R}(s) = \mathbb{E}_{X\sim\mu,\Sigma\sim P_X}\left[d_\tau\left(s(X),\Sigma\right)\right], \tag{1}$$

where d is the Kendall's tau distance, i.e for all $(\sigma,\sigma') \in \mathfrak{S}_n^2$:

$$d_\tau(\sigma,\sigma') = \sum_{i<j} \mathbb{I}\{(\sigma(j)-\sigma(i))\cdot(\sigma'(j)-\sigma'(i)) < 0\}$$

Definition

For any $\sigma \in \mathfrak{S}_n$, the risk of σ is $L(\sigma) = \mathbb{E}_{\Sigma\sim P}[d(\Sigma,\sigma)]$. A **Kemeny median** is any

$$\sigma_P^* = \operatorname*{argmin}_{\sigma\in\mathfrak{S}_n} L(\sigma) \tag{2}$$

Optimal predictors s^* minimizing (1) are the rules that maps any point X to any Kemeny median of P_X (minimizing (2)):

$$\boxed{s^* = \operatorname*{argmin}_{s\in\mathcal{S}} \mathcal{R}(s) \quad \Leftrightarrow \quad s(X) = \sigma_{P_X}^*}$$

Local Consensus Methods for Ranking Median Regressions

Idea: approximate s^* with piecewise constant ranking rules ($s \in \mathcal{S}_{\mathcal{P}}$), by computing **local Kemeny medians**.

Local Consensus Methods for Ranking Median Regressions

Idea: approximate s^* with piecewise constant ranking rules ($s \in \mathcal{S}_{\mathcal{P}}$), by computing **local Kemeny medians**.

Empirical Risk Minimization: Consider a statistical version of theoretical risk based on the training data (X_i, Σ_i)'s:

$$\widehat{\mathcal{R}}_N(s) = \frac{1}{N}\sum_{k=1}^{N} d_\tau(s(X_k),\ \Sigma_k)$$

and solutions of the optimization problem:

$$\min_{s \in \mathcal{S}_{\mathcal{P}}} \widehat{\mathcal{R}}_N(s),$$

Local Consensus Methods for Ranking Median Regressions

Idea: approximate s^* with piecewise constant ranking rules ($s \in \mathcal{S}_{\mathcal{P}}$), by computing **local Kemeny medians**.

Empirical Risk Minimization: Consider a statistical version of theoretical risk based on the training data (X_i, Σ_i)'s:

$$\widehat{\mathcal{R}}_N(s) = \frac{1}{N}\sum_{k=1}^{N} d_\tau(s(X_k),\ \Sigma_k)$$

and solutions of the optimization problem:

$$\min_{s \in \mathcal{S}_{\mathcal{P}}} \widehat{\mathcal{R}}_N(s),$$

Our Approach

Generate partitions tailored to the training data and yielding a ranking rule with nearly minimum predictive error. Two methods are investigated: K-Nearest Neighbor and Decision-tree.

Thank you!