Supplement to "The Kendall and Mallows Kernels for Permutations"

Yunlong Jiao YUNLONG.JIAO@MINES-PARISTECH.FR
Jean-Philippe Vert JEAN-PHILIPPE.VERT@MINES-PARISTECH.FR
MINES ParisTech – CBIO, PSL Research University, Institut Curie, INSERM U900, Paris, France

Abstract

This document contains proofs and algorithms to supplement the paper titled "The Kendall and Mallows kernels for permutations" accepted to ICML 2015.

1. Proof to Theorem 3

Proof. Let $\widehat{\mathbf{w}}$ be a solution to the original SVM optimization problem, and $\widehat{\mathbf{w}}_D$ a solution to the perturbed SVM, i.e., a solution of

$$\min_{\mathbf{w}} F_D(\mathbf{w}) = \frac{\lambda}{2}\|\mathbf{w}\|^2 + \widehat{R}_D(\mathbf{w}), \tag{1}$$

with $\widehat{R}_D(\mathbf{w}) = \frac{1}{m}\sum_{i=1}^m \ell(y_i \mathbf{w}^\top \Psi_D(\mathbf{x}_i))$. Since the hinge loss is 1-Lipschitz, i.e., $|\ell(a) - \ell(b)| \leq |a-b|$ for any $a, b \in \mathbb{R}$, we obtain that for any $\mathbf{u} \in \mathbb{R}^{\binom{n}{2}}$:

$$\begin{aligned}\left| \widehat{R}(\mathbf{u}) - \widehat{R}_D(\mathbf{u}) \right| &\leq \frac{1}{m}\sum_{i=1}^m \left| \mathbf{u}^\top \left(\Psi(\mathbf{x}_i) - \Psi_D(\mathbf{x}_i)\right) \right| \\ &\leq \| \mathbf{u} \| \sup_{i=1,\dots,m} \| \Psi_D(\mathbf{x}_i) - \Psi(\mathbf{x}_i) \| .\end{aligned} \tag{2}$$

Now, since $\widehat{\mathbf{w}}_D$ is a solution of (1), it satisfies

$$\| \widehat{\mathbf{w}}_D \| \leq \sqrt{\frac{2F_D(\widehat{\mathbf{w}}_D)}{\lambda}} \leq \sqrt{\frac{2F_D(0)}{\lambda}} = \sqrt{\frac{2}{\lambda}} ,$$

and similarly $\| \widehat{\mathbf{w}} \| \leq \sqrt{2/\lambda}$ because $\widehat{\mathbf{w}}$ is a solution of the original SVM optimization problem. Using (2) and these bounds on $\| \widehat{\mathbf{w}}_D \|$ and $\| \widehat{\mathbf{w}} \|$, we get

$$\begin{aligned} & F(\widehat{\mathbf{w}}_D) - F(\widehat{\mathbf{w}}) \\ &\quad = F(\widehat{\mathbf{w}}_D) - F_D(\widehat{\mathbf{w}}_D) + F_D(\widehat{\mathbf{w}}_D) - F(\widehat{\mathbf{w}}) \\ &\quad \leq F(\widehat{\mathbf{w}}_D) - F_D(\widehat{\mathbf{w}}_D) + F_D(\widehat{\mathbf{w}}) - F(\widehat{\mathbf{w}}) \\ &\quad = \widehat{R}(\widehat{\mathbf{w}}_D) - \widehat{R}_D(\widehat{\mathbf{w}}_D) + \widehat{R}_D(\widehat{\mathbf{w}}) - \widehat{R}(\widehat{\mathbf{w}}) \\ &\quad \leq (\| \widehat{\mathbf{w}}_D \| + \| \widehat{\mathbf{w}} \|) \sup_{i=1,\dots,m} \| \Psi_D(\mathbf{x}_i) - \Psi(\mathbf{x}_i) \| \\ &\quad \leq \sqrt{\frac{8}{\lambda}} \sup_{i=1,\dots,m} \| \Psi_D(\mathbf{x}_i) - \Psi(\mathbf{x}_i) \| .\end{aligned} \tag{3}$$

Theorem 3 then follows from the following lemma. □

Lemma 1. *For any $0 < \delta < 1$, the following holds with probability greater than $1 - \delta$:*

$$\sup_{i=1,\dots,m} \| \Psi_D(\mathbf{x}_i) - \Psi(\mathbf{x}_i) \| \leq \frac{1}{\sqrt{D}} \left(2 + \sqrt{8 \log \frac{m}{\delta}} \right) .$$

Proof. For any $i \in [1, m]$, we can apply Boucheron et al. (2013, Example 6.3) to the random vector $X_j = \Phi(\tilde{\mathbf{x}}_i^j) - \Psi(\mathbf{x}_i)$ that satisfies $\mathbb{E}X_j = 0$ and $\| X_j \| \leq 2$ a.s. to get, for any $u \geq 2/\sqrt{D}$,

$$\mathbb{P}\left(\| \Psi_D(\mathbf{x}_i) - \Psi(\mathbf{x}_i) \| \geq u \right) \leq \exp\left(- \frac{\left(u\sqrt{D} - 2 \right)^2}{8} \right) .$$

Lemma 1 then follows by a simple union bound. □

References

Boucheron, S., Lugosi, G., and Massart, P. *Concentration Inequalities*. Oxford Univ Press, 2013.

Proceedings of the 32^{nd} International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37.

Algorithm 1 Kendall kernel for two interleaving partial rankings.

Input: two partial rankings $A_{i_1,\ldots,i_k}, A_{j_1,\ldots,j_m} \subset \mathbb{S}_n$, corresponding to subsets of item indices $I := \{i_1, \ldots, i_k\}$ and $J := \{j_1, \ldots, j_m\}$.

1: Let $\sigma \in \mathbb{S}_k$ be the total ranking corresponding to the k observed items in $A_{i_1,\ldots,i_k}$, and $\sigma' \in \mathbb{S}_m$ be the total ranking corresponding to the m observed items in $A_{j_1,\ldots,j_m}$.

2: Let $\tau \in \mathbb{S}_{|I\cap J|}$ be the total ranking of the observed items indexed by $I \cap J$ in $A_{i_1,\ldots,i_k}$, and $\tau' \in \mathbb{S}_{|I\cap J|}$ the total ranking of the observed items indexed by $I \cap J$ in partial ranking $A_{j_1,\ldots,j_m}$.

3: Initialize $s_1 = s_2 = s_3 = s_4 = s_5 = 0$.

4: If $|I \cap J| \geq 2$, update

$$s_1 = \frac{\binom{|I\cap J|}{2}}{\binom{n}{2}} K(\tau, \tau').$$

5: If $|I \cap J| \geq 1$ and $|I \setminus J| \geq 1$, update

$$s_2 = \frac{1}{\binom{n}{2}(m+1)} \sum_{l \in I\cap J} \Big\{ \big[2\sigma'(l) - m - 1\big] \times \big[2(\sigma(l) - \tau(l)) - k + |I \cap J|\big] \Big\}.$$

6: If $|I \cap J| \geq 1$ and $|J \setminus I| \geq 1$, update

$$s_3 = \frac{1}{\binom{n}{2}(k+1)} \sum_{l \in I\cap J} \Big\{ \big[2\sigma(l) - k - 1\big] \times \big[2(\sigma'(l) - \tau'(l)) - m + |I \cap J|\big] \Big\}.$$

7: If $|I \cap J| \geq 1$ and $|(I \cup J)^{\complement}| \geq 1$, update

$$s_4 = \frac{|(I \cup J)^{\complement}|}{\binom{n}{2}(k+1)(m+1)} \times \sum_{l \in I\cap J} \big[2\sigma(l) - k - 1\big]\big[2\sigma'(l) - m - 1\big].$$

8: If $|I \setminus J| \geq 1$ and $|J \setminus I| \geq 1$, update

$$s_5 = \frac{-1}{\binom{n}{2}(k+1)(m+1)} \times \sum_{l \in I\setminus J} \big[2\sigma(l) - k - 1\big] \sum_{v \in J\setminus I} \big[2\sigma'(v) - m - 1\big].$$

Output: $K(A_{i_1,\ldots,i_k}, A_{j_1,\ldots,j_m}) = s_1 + s_2 + s_3 + s_4 + s_5$.

Algorithm 2 Kendall kernel for a top-k partial ranking and a top-m partial ranking.

Input: a top-k partial ranking and a top-m partial ranking $B_{i_1,\ldots,i_k}, B_{j_1,\ldots,j_m} \subset \mathbb{S}_n$, corresponding to subsets of item indices $I := \{i_1, \ldots, i_k\}$ and $J := \{j_1, \ldots, j_m\}$.

1: Let $\sigma \in \mathbb{S}_k$ be the total ranking corresponding to the k observed items in $B_{i_1,\ldots,i_k}$, and $\sigma' \in \mathbb{S}_m$ be the total ranking corresponding to the m observed items in $B_{j_1,\ldots,j_m}$.

2: Let $\tau \in \mathbb{S}_{|I\cap J|}$ be the total ranking of the observed items indexed by $I \cap J$ in $B_{i_1,\ldots,i_k}$, and $\tau' \in \mathbb{S}_{|I\cap J|}$ the total ranking of the observed items indexed by $I \cap J$ in partial ranking $B_{j_1,\ldots,j_m}$.

3: Initialize $s_1 = s_2 = s_3 = s_4 = s_5 = 0$.

4: If $|I \cap J| \geq 2$, update

$$s_1 = \frac{\binom{|I\cap J|}{2}}{\binom{n}{2}} K(\tau, \tau').$$

5: If $|I \cap J| \geq 1$ and $|I \setminus J| \geq 1$, update

$$s_2 = \frac{1}{\binom{n}{2}} \sum_{l \in I\cap J} \big[2(\sigma(l) - \tau(l)) - k + |I \cap J|\big].$$

6: If $|I \cap J| \geq 1$ and $|J \setminus I| \geq 1$, update

$$s_3 = \frac{1}{\binom{n}{2}} \sum_{l \in I\cap J} \big[2(\sigma'(l) - \tau'(l)) - m + |I \cap J|\big].$$

7: If $|I \cap J| \geq 1$ and $|(I \cup J)^{\complement}| \geq 1$, update

$$s_4 = \frac{|I \cap J| \cdot |(I \cup J)^{\complement}|}{\binom{n}{2}}.$$

8: If $|I \setminus J| \geq 1$ and $|J \setminus I| \geq 1$, update

$$s_5 = \frac{-|I \setminus J| \cdot |J \setminus I|}{\binom{n}{2}}.$$

Output: $K(B_{i_1,\ldots,i_k}, B_{j_1,\ldots,j_m}) = s_1 + s_2 + s_3 + s_4 + s_5$.