The weighted Kendall and high-order kernels for permutations

Yunlong Jiao
University of Oxford

Jean-Philippe Vert
MINES ParisTech

ICML @ Stockholm, 11 July 2018

Ranking data is everywhere

- Rank data, e.g., preference survey:

- Ranking extracted from data, e.g., gene expression:

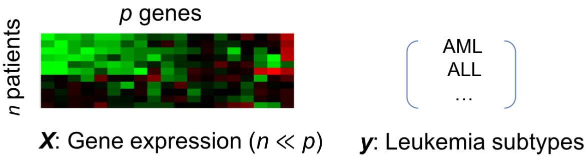
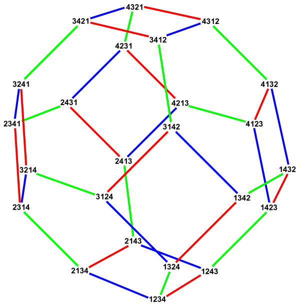


Figure: "If the expression of two genes SPTAN1 $\geq$ CD33, then ALL; otherwise AML" gives accuracy of 93.80% [Tan et al., 2005].

Mathematical difficulty



- A (total) ranking is a permutation

$$\sigma : [1, n] \to [1, n]$$

  such that $\sigma(i)$ = rank of item i.

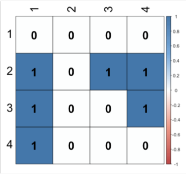
- $\mathbb{S}_n$ the symmetric group with composition

$$(\sigma_1 \sigma_2)(i) = \sigma_1(\sigma_2(i))$$

  with cardinality $|\mathbb{S}_n| = n!$.

The Kendall embedding

i	1	2	3	4
$\sigma(i)$	4	1	2	3



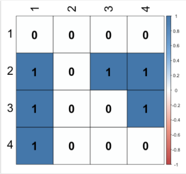
$\Phi_{ij}(\sigma)$

- The Kendall embedding is an n^2-dimensional Euclidean embedding:

$$\Phi_\tau : \mathbb{S}_n \to \mathbb{R}^{n\times n} ; \quad (\Phi_\tau(\sigma))_{i,j} = \begin{cases} 1 & \text{if } \sigma(i) < \sigma(j) \,, \\ 0 & \text{otherwise.} \end{cases}$$

The Kendall embedding

i	1	2	3	4
$\sigma(i)$	4	1	2	3



$\Phi_{ij}(\sigma)$

- Quantities using Kendall embedding [Jiao and Vert, 2015]:

Kendall's τ correlation: $K_\tau(\sigma, \sigma') \propto \langle \Phi_\tau(\sigma), \Phi_\tau(\sigma') \rangle$

Kendall's τ distance: $d_\tau(\sigma, \sigma') \propto \|\Phi_\tau(\sigma) - \Phi_\tau(\sigma')\|^2$

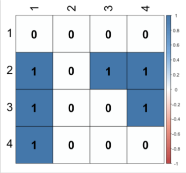
Mallows distribution: $\mathbb{P}(\sigma) \propto \exp\left\{-\lambda \|\Phi_\tau(\sigma) - \Phi_\tau(\pi)\|^2\right\}$
for mode $\pi \in \mathbb{S}_n$ and dispersion $\lambda > 0$

(Nonlinear) predictive model: $h(\sigma) = \langle B, \Phi_\tau(\sigma) \rangle$
for coefficients $B \in \mathbb{R}^{n \times n}$

$\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denote the Frobenius inner product and norm.

The Kendall kernel

i	1	2	3	4
$\sigma(i)$	4	1	2	3



$\Phi_{ij}(\sigma)$

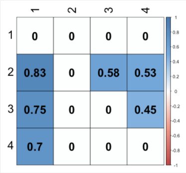
- Quantities using Kendall embedding [Jiao and Vert, 2015]:

Kendall kernel:	$K_\tau(\sigma,\sigma') \propto \langle \Phi_\tau(\sigma), \Phi_\tau(\sigma') \rangle$
Kendall's τ distance:	$d_\tau(\sigma,\sigma') \propto \|\Phi_\tau(\sigma) - \Phi_\tau(\sigma')\|^2$
Mallows distribution:	$\mathbb{P}(\sigma) \propto \exp\left\{-\lambda\|\Phi_\tau(\sigma) - \Phi_\tau(\pi)\|^2\right\}$ for mode $\pi \in \mathbb{S}_n$ and dispersion $\lambda > 0$
Kernel machine (Classification, regression, clustering, etc.):	$h(\sigma) = \sum_{i=1}^N \alpha_i K_\tau(\sigma,\sigma_i)$ for coefficients $\alpha \in \mathbb{R}^N$ ("kernel trick")

$\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ denote the Frobenius inner product and norm.

The weighted Kendall embedding

i	1	2	3	4
$\sigma(i)$	4	1	2	3



$\Phi^U{}_{ij}(\sigma)$

- The weighted Kendall embedding is:

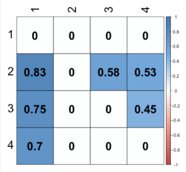
$$\Phi^U : \mathbb{S}_n \to \mathbb{R}^{n\times n}\ ; \quad \left(\Phi^U(\sigma)\right)_{i,j} = \begin{cases} U_{\sigma(i),\sigma(j)} & \text{if } \sigma(i) < \sigma(j)\,, \\ 0 & \text{otherwise,} \end{cases}$$

where $U \in \mathbb{R}^{n\times n}$ is a weight matrix such as

- Top-k $U_{i,j} = 1$ iff $i, j \leq k$, for rank threshold k.
- Additive $U_{i,j} = u_i + u_j$, for rank discounts $u \in \mathbb{R}^n$.
- Multiplicative $U_{i,j} = u_i u_j$, for rank discounts $u \in \mathbb{R}^n$.

The weighted Kendall kernel

i	1	2	3	4
$\sigma(i)$	4	1	2	3



$\Phi^{U}_{ij}(\sigma)$

- The weighted Kendall kernel:

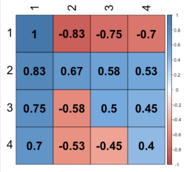
$$K_U(\sigma, \sigma') \propto \left\langle \Phi^U(\sigma), \Phi^U(\sigma') \right\rangle$$

is a weighted extension of the Kendall's tau correlation, and we showed that it is:

- ✓ Symmetric.
- ✓ Invariant to shuffling of the index i.
- ✓ Fast to compute in $O(n \log n)$ time.
- ✓ Positive definite, enabling to use all kernel machines (classification, regression, clustering, etc.)

Learning the weights

i	1	2	3	4
$\sigma(i)$	4	1	2	3



$\Psi^U{}_{ij}(\sigma)$

- In general, the weighted embedding and kernel are:

$$\Phi^U : \mathbb{S}_n \to \mathbb{R}^{n\times n}\,; \quad \left(\Phi^U(\sigma)\right)_{i,j} = U_{\sigma(i),\sigma(j)}\,,$$

$$G_U(\sigma,\sigma') \propto \left\langle \Phi^U(\sigma), \Phi^U(\sigma') \right\rangle\,.$$

- The weight matrix $U \in \mathbb{R}^{n\times n}$ can be learned jointly with the coefficients of a kernel machine, by solving a non-convex optimization via:
  - Alternative optimization between weights and coefficients.
  - Low-rank approximation [Le Morvan and Vert, 2017].

Going beyond item pairs

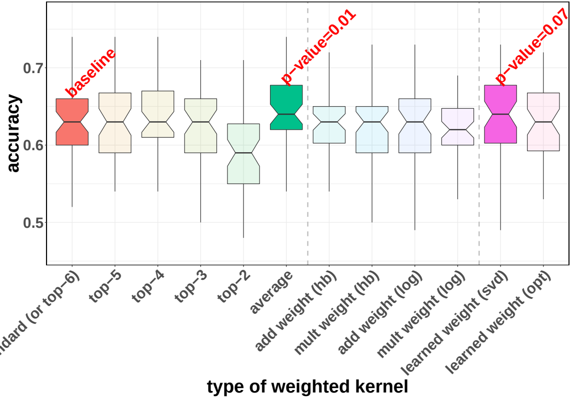
- The order-3 (or higher) weighted embedding and kernel are:

$$\Phi^{\mathcal{U}} : \mathbb{S}_n \to \mathbb{R}^{n \times n \times n} \,; \quad \left(\Phi^{\mathcal{U}}(\sigma)\right)_{i,j,k} = \mathcal{U}_{\sigma(i),\sigma(j),\sigma(k)} \,,$$
$$G_{\mathcal{U}}(\sigma, \sigma') \propto \left\langle \Phi^{\mathcal{U}}(\sigma), \Phi^{\mathcal{U}}(\sigma') \right\rangle \,.$$

- The weight tensor $\mathcal{U} \in \mathbb{R}^{n \times n \times n}$ can also be learned in a data-driven way, similarly to the order-2 case.

Real-data experiments: Prediction accuracy

- Eurobarometer survey data [Christensen, 2010]:
  - $>$12k participants ranked the importance of 6 sources of information
  - Binary classification problem: predict age group ($>$40yo vs $<$40yo) from ranking



Real-data experiments: Weights learned

- Eurobarometer survey data [Christensen, 2010]:
  - >12k participants ranked the importance of 6 sources of information
  - Binary classification problem: predict age group (>40yo vs <40yo) from ranking

Conclusion

- We have studied
  - ✓ Euclidean embeddings and positive definite kernels

  for permutations, which are
  - ✓ weighted and high-order extensions

  to Kendall's tau correlation, where
  - ✓ weights can be learned systematically in a data-driven way.

References I

- Pictures are downloaded from internet (no copyright infringement is intended). https://www.dreamstime.com/stock-illustration-chart-showing-rank-people-lives-most-successful-one-top-here-illustration-image53416790; https://www.dreamstime.com/stock-photos-speed-limit-35-image754943.
- Christensen, T. (2010).
  Eurobarometer 55.2: Science and technology, agriculture, the euro, and internet access, may-june 2001.
  https://doi.org/10.3886/ICPSR03341.v3.
- Jiao, Y. and Vert, J.-P. (2015).
  The Kendall and Mallows kernels for permutations.
  In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, volume 37 of *Proceedings of Machine Learning Research (PMLR)*, pages 1935–1944.
- Le Morvan, M. and Vert, J.-P. (2017).
  Supervised quantile normalisation.
  arXiv preprint arXiv:1706.00244.
- Tan, A. C., Naiman, D. Q., Xu, L., Winslow, R. L., and Geman, D. (2005).
  Simple decision rules for classifying human cancers from gene expression profiles.
  Bioinformatics, 21(20):3896–3904.