Poster Session 6E Tonight!

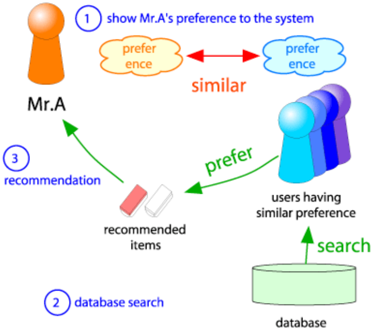
The Kendall and Mallows Kernels for Permutations

Yunlong Jiao & Jean-Philippe Vert

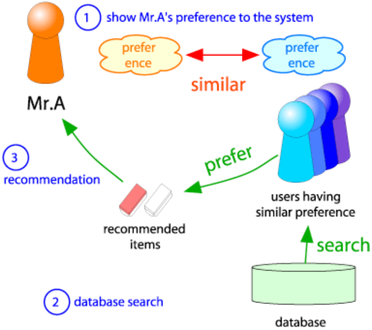
MINES ParisTech

ICML Lille, July 8, 2015

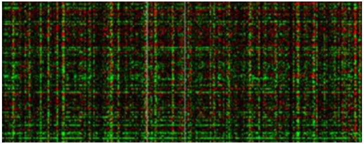
- Recommender system, e.g. Collaborative Filtering.



- Recommender system, e.g. Collaborative Filtering.



- Learn from converted rankings, e.g., gene expression data analysis for leukemia classification [Tan et al., 2005].
  - Data: $n \times p$ matrix ($p \gg n$).



  - Rule: if SPTAN1 $\geq$ CD33 then ALL; else AML.
  - Accuracy: 93.80% (LOOCV).

Outline

1. Data type
   - Rankings and permutations.
2. Methods
   - Computationally efficient kernels for total rankings, partial rankings and rankings converted from quantitative vectors.
3. Experiments
   - High-dimensional classification in biomedical applications.

Total Rankings and Permutations

- A total ranking is a strict ordering of n items $\{x_1, x_2, \ldots, x_n\}$,

$$x_{i_1} \succ x_{i_2} \succ \cdots \succ x_{i_n}\,.$$

- A permutation is a rearrangement of n indices,

  $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ such that $\sigma(i) \neq \sigma(j)$ for $i \neq j$.

Total Rankings and Permutations

- A total ranking is a strict ordering of n items $\{x_1, x_2, \ldots, x_n\}$,

$$x_{i_1} \succ x_{i_2} \succ \cdots \succ x_{i_n}\,.$$

- A permutation is a rearrangement of n indices,

$\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ such that $\sigma(i) \neq \sigma(j)$ for $i \neq j$.

- A total ranking is **equivalently represented** by a permutation if σ maps item index to item rank, e.g.,

$$\begin{aligned} & x_2 \succ x_4 \succ x_3 \succ x_1 \\ \iff & \sigma = \begin{pmatrix} 2 & 4 & 3 & 1 \\ 4 & 3 & 2 & 1 \end{pmatrix} \begin{matrix} \text{— index} \\ \text{— rank} \end{matrix} \\ \iff & \sigma(1) = 1, \sigma(2) = 4, \sigma(3) = 2, \sigma(4) = 3\,. \end{aligned}$$

Kendall tau Distance for Permutations

- Kendall tau distance [Kendall, 1938] counts the number of discordant pairs between permutations, i.e.,

$$n_d(\sigma, \sigma') = \sum_{i<j} \mathbb{1}_{\sigma(i)<\sigma(j)} \mathbb{1}_{\sigma'(i)>\sigma'(j)} + \mathbb{1}_{\sigma(i)>\sigma(j)} \mathbb{1}_{\sigma'(i)<\sigma'(j)} .$$

E.g.,

index e	1	2	3	4
rank σ	2	3	4	1
rank σ'	3	1	4	2

$$n_d(\sigma, \sigma') = 1 + 1 + 0 = 2$$

Kendall tau Distance for Permutations

- Kendall tau distance [Kendall, 1938] counts the number of discordant pairs between permutations, i.e.,

$$n_d(\sigma, \sigma') = \sum_{i<j} \mathbb{1}_{\sigma(i)<\sigma(j)} \mathbb{1}_{\sigma'(i)>\sigma'(j)} + \mathbb{1}_{\sigma(i)>\sigma(j)} \mathbb{1}_{\sigma'(i)<\sigma'(j)} .$$

- The number of concordant pairs between permutations is

$$n_c(\sigma, \sigma') = \binom{n}{2} - n_d(\sigma, \sigma') .$$

E.g.,

index e	1	2	3	4
rank σ	2	3	4	1
rank σ'	3	1	4	2

$$n_d(\sigma, \sigma') = 1 + 1 + 0 = 2$$

$$n_c(\sigma, \sigma') = \frac{4(4-1)}{2} - 2 = 4$$

Kendall and Mallows Kernels for Permutations

- The Kendall tau coefficient is defined as

$$K_\tau(\sigma, \sigma') = \frac{n_c(\sigma, \sigma') - n_d(\sigma, \sigma')}{\binom{n}{2}}\,.$$

- The Mallows measure is defined for any $\lambda \geq 0$ by

$$K_M^\lambda(\sigma, \sigma') = e^{-\lambda n_d(\sigma, \sigma')}\,.$$

E.g.,

index e	1	2	3	4
rank σ	2	3	4	1
rank σ'	3	1	4	2

$$K_\tau(\sigma, \sigma') = \frac{4-2}{6} = \frac{1}{3}$$

$$K_M^\lambda(\sigma, \sigma') = e^{-2\lambda}, \lambda \geq 0$$

Kendall and Mallows Kernels for Permutations

- The Kendall tau coefficient is defined as

$$K_\tau(\sigma,\sigma') = \frac{n_c(\sigma,\sigma') - n_d(\sigma,\sigma')}{\binom{n}{2}}.$$

- The Mallows measure is defined for any $\lambda \geq 0$ by

$$K_M^\lambda(\sigma,\sigma') = e^{-\lambda n_d(\sigma,\sigma')}.$$

E.g.,

index e	1	2	3	4
rank σ	2	3	4	1
rank σ'	3	1	4	2

$$K_\tau(\sigma,\sigma') = \frac{4-2}{6} = \frac{1}{3}$$

$$K_M^\lambda(\sigma,\sigma') = e^{-2\lambda}, \lambda \geq 0$$

Theorem (Main theorem)

These two similarity measures for permutations are positive definite kernels.

Kendall and Mallows Kernels for Permutations

- The Kendall kernel is defined as

$$K_\tau(\sigma, \sigma') = \frac{n_c(\sigma, \sigma') - n_d(\sigma, \sigma')}{\binom{n}{2}} .$$

- The Mallows kernel is defined for any $\lambda \geq 0$ by

$$K_M^\lambda(\sigma, \sigma') = e^{-\lambda n_d(\sigma, \sigma')} .$$

Theorem (Main theorem)

These two kernels for permutations are positive definite.

Proof.

Consider the explicit kernel mapping

$$\Phi : \mathbb{S}_n \to \mathbb{R}^{\binom{n}{2}}, \sigma \mapsto \Big(sgn(\sigma(i) - \sigma(j)) \Big)_{1 \leq i < j \leq n} .$$

The Kendall and Mallows kernel correspond respectively to a linear and Gaussian kernel on a $\binom{n}{2}$-dimensional embedding of $\mathbb{S}_n$. □

Kendall and Mallows Kernels for Permutations

- The Kendall kernel is defined as

$$K_\tau(\sigma, \sigma') = \frac{n_c(\sigma, \sigma') - n_d(\sigma, \sigma')}{\binom{n}{2}} \,.$$

- The Mallows kernel is defined for any $\lambda \geq 0$ by

$$K_M^\lambda(\sigma, \sigma') = e^{-\lambda n_d(\sigma, \sigma')} \,.$$

Theorem (Main theorem)

These two kernels for permutations are positive definite.

Theorem ([Knight, 1966])

These two kernels for permutations can be evaluated in $O(n \log n)$ *time.*

- Two interesting types of partial rankings are interleaving partial ranking

$$x_{i_1} \succ x_{i_2} \succ \cdots \succ x_{i_k}, \quad k \leq n.$$

and top-k partial ranking

$$x_{i_1} \succ x_{i_2} \succ \cdots \succ x_{i_k} \succ X_{\mathrm{rest}}, \quad k \leq n.$$

- Partial rankings can be **uniquely represented** by a set of permutations compatible with all the observed partial orders.

Convolution Kendall Kernel for Partial Rankings

- Two interesting types of partial rankings are interleaving partial ranking

$$x_{i_1} \succ x_{i_2} \succ \cdots \succ x_{i_k}, \quad k \leq n.$$

and top-k partial ranking

$$x_{i_1} \succ x_{i_2} \succ \cdots \succ x_{i_k} \succ X_{\mathrm{rest}}, \quad k \leq n.$$

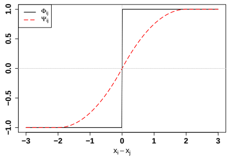
- Partial rankings can be **uniquely represented** by a set of permutations compatible with all the observed partial orders.

Theorem

For these two particular types of partial rankings, the convolution kernel [Haussler, 1999] induced by Kendall kernel

$$K_\tau^\star(R, R') = \frac{1}{|R||R'|} \sum_{\sigma \in R} \sum_{\sigma' \in R'} K_\tau(\sigma, \sigma')$$

can be evaluated in $O(k \log k)$ *time.*



- Kendall mapping for quantitative vectors is discrete-valued and very sensitive to "almost ties", i.e.,

$$\Phi : \mathbb{R}^n \to \mathbb{R}^{\binom{n}{2}}, \mathbf{x} \mapsto \left(\mathbb{1}_{x_i > x_j} - \mathbb{1}_{x_i < x_j} \right)_{1 \le i < j \le n} .$$

- We propose a noise-corrupted kernel mapping instead (similarly to [Muandet et al., 2012])

$$\Psi(\mathbf{x}) = \mathbb{E}\Phi(\underbrace{\mathbf{x} + \epsilon}_{\tilde{\mathbf{x}}}) = \Big(\mathbb{P}\left(\tilde{x}_i > \tilde{x}_j\right) - \mathbb{P}\left(\tilde{x}_i < \tilde{x}_j\right) \Big)_{1 \le i < j \le n} .$$

- Kendall kernel stabilized alternative is given by

$$G\left(\mathbf{x}, \mathbf{x}'\right) = \Psi(\mathbf{x})^\top \Psi(\mathbf{x}') = \mathbb{E} K_\tau(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}') .$$

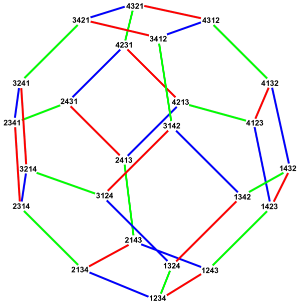


Figure : Cayley graph of $\mathbb{S}_4$.

- Diffusion kernel [Kondor and Lafferty, 2002] is defined by

$$K^{\beta}_{\mathrm{dif}}(\sigma, \sigma') = [e^{\beta\Delta}]_{\sigma,\sigma'} ,$$

where Δ is the graph laplacian.

- Mallows kernel is written as

$$K^{\lambda}_{M}(\sigma, \sigma') = e^{-\lambda n_d(\sigma,\sigma')} ,$$

where $n_d(\sigma, \sigma') = d_{\mathcal{G}}(\sigma, \sigma')$ the shortest path distance on graph.

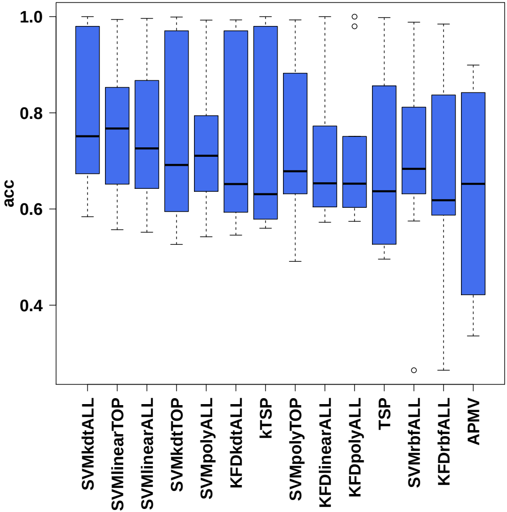
Datasets

Dataset	No. of features	No. of samples (training/test) C_1	C_2
Breast Cancer 1	23624	44/7 (Non-relapse)	32/12 (Relapse)
Breast Cancer 2	22283	142 (Non-relapse)	56 (Relapse)
Breast Cancer 3	22283	71 (Poor Prognosis)	138 (Good Prognosis)
Colon Tumor	2000	40 (Tumor)	22 (Normal)
Lung Cancer 1	7129	24 (Poor Prognosis)	62 (Good Prognosis)
Lung Cancer 2	12533	16/134 (ADCA)	16/15 (MPM)
Medulloblastoma	7129	39 (Failure)	21 (Survivor)
Ovarian Cancer	15154	162 (Cancer)	91 (Normal)
Prostate Cancer 1	12600	50/9 (Normal)	52/25 (Tumor)
Prostate Cancer 2	12600	13 (Non-relapse)	8 (Relapse)

Methods

- Kernel machines Support Vector Machines (SVM) and Kernel Fisher Discriminant (KFD) with Kendall kernel, linear kernel, Gaussian RBF kernel, polynomial kernel.
- Top Scoring Pairs (TSP) classifiers [Tan et al., 2005].
- Hybrid scheme of SVM + TSP feature selection algorithm.

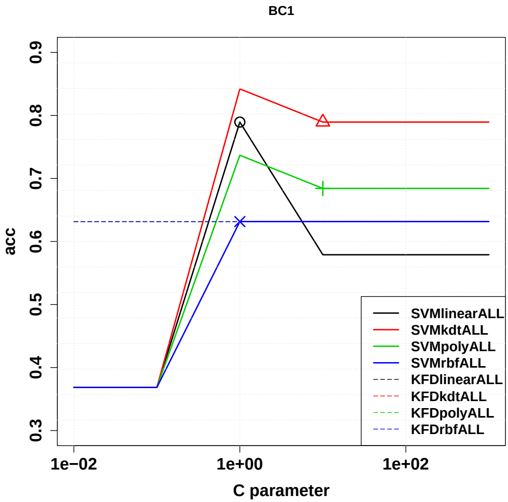
Results



Kendall kernel SVM

- **Competitive accuracy!**

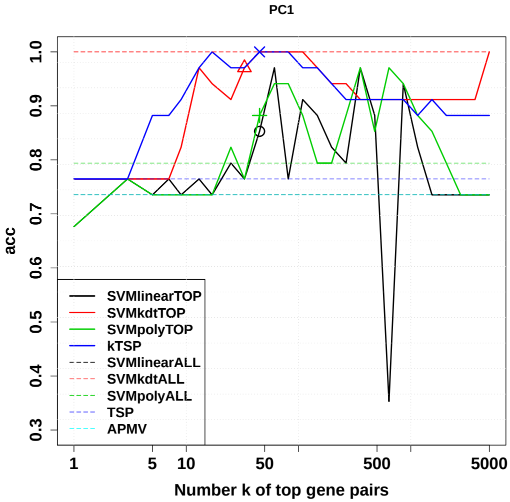
Results



Kendall kernel SVM

- Competitive accuracy!
- **Insensitive to C parameter!**

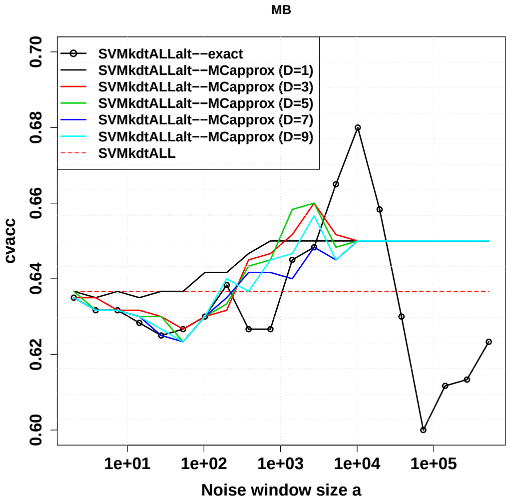
Results



Kendall kernel SVM

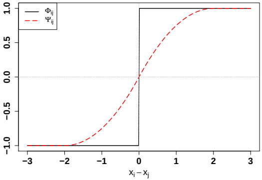
- Competitive accuracy!
- Insensitive to C parameter!
- **No feature selection!**

Results



Kendall kernel Support Measure Machines [Muandet et al., 2012]

- **Improved accuracy!**



Conclusion

Poster Session 6E Tonight!

IF you are dealing with ranking-related problems,

IF your problem can be formulated in a way that some kernel machine can cope with,

DO throw Kendall and Mallows kernel into that kernel machine!

Acknowledgments

This work was supported by the European Union 7th Framework Program through the Marie Curie ITN MLPM grant No 316861, and by the European Research Council grant ERC-SMAC-280032.

Machine Learning for Personalized Medicine

European Commission

References I

Haussler, D. (1999).
Convolution kernels on discrete structures.
Technical Report UCSC-CRL-99-10, UC Santa Cruz.

Kendall, M. G. (1938).
A new measure of rank correlation.
Biometrika, 30(1/2):81–93.

Knight, W. R. (1966).
A computer method for calculating Kendall's tau with ungrouped data.
J. Am. Stat. Assoc., 61(314):436–439.

Kondor, I. R. and Lafferty, J. (2002).
Diffusion kernels on graphs and other discrete input spaces.
In *Proceedings of the Nineteenth International Conference on Machine Learning*, volume 2, pages 315–322, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Muandet, K., Fukumizu, K., Dinuzzo, F., and Schölkopf, B. (2012).
Learning from distributions via support measure machines.
In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Adv. Neural. Inform. Process Syst.*, volume 25, pages 10–18. Curran Associates, Inc.

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.