Bayesian Efficient Multiple Kernel Learning

Mehmet Gönen mehmet.gonen@aalto.fi http://users.ics.aalto.fi/gonen

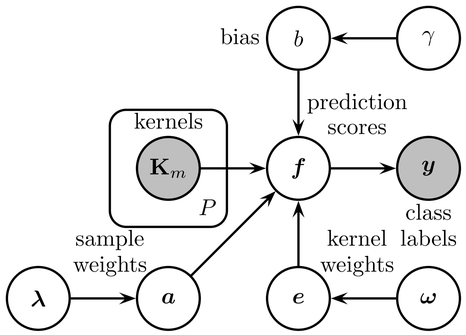
Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University

Motivation

- To obtain a better similarity measure and to integrate information from different sources
- To develop a computationally feasible Bayesian MKL algorithm **without sampling**

Existing Bayesian Methods

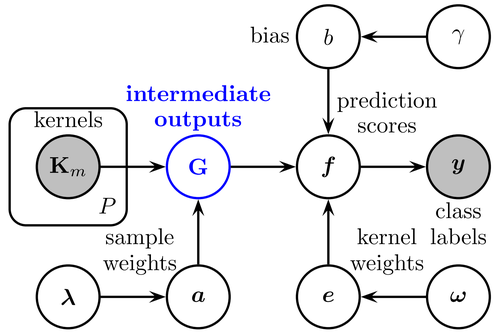
- *Nonconjugacy* between Dirichlet and normal distributions requires a sampling method
- *Nonlinear dependency* between random variables when calculating prediction scores



Proposed Method

- Combination is formulated in a novel way
- **Intermediate outputs** are introduced as auxiliary variables
- Kernel weights are assumed to be **normally distributed** without any constraints
- **Sample- and kernel-level sparsities** can be adjusted using gamma priors on precisions

Graphical Model



Probabilistic Model

$$
\begin{aligned}
\lambda_i &\sim \mathcal{G}(\lambda_i; \alpha_\lambda, \beta_\lambda) && \forall i \\
a_i|\lambda_i &\sim \mathcal{N}(a_i; 0, \lambda_i^{-1}) && \forall i \\
g_i^m|\boldsymbol{a}, \boldsymbol{k}_{m,i} &\sim \mathcal{N}(g_i^m; \boldsymbol{a}^\top \boldsymbol{k}_{m,i}, 1) && \forall (m,i) \\
\gamma &\sim \mathcal{G}(\gamma; \alpha_\gamma, \beta_\gamma) \\
b|\gamma &\sim \mathcal{N}(b; 0, \gamma^{-1}) \\
\omega_m &\sim \mathcal{G}(\omega_m; \alpha_\omega, \beta_\omega) && \forall m \\
e_m|\omega_m &\sim \mathcal{N}(e_m; 0, \omega_m^{-1}) && \forall m \\
f_i|b, \boldsymbol{e}, \boldsymbol{g}_i &\sim \mathcal{N}(f_i; \boldsymbol{e}^\top \boldsymbol{g}_i + b, 1) && \forall i \\
y_i|f_i &\sim \delta(f_i y_i > \nu) && \forall i
\end{aligned}
$$

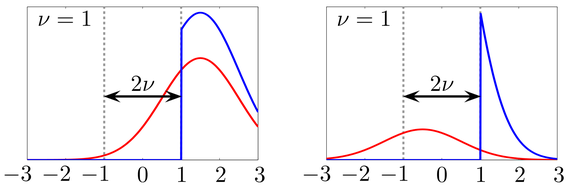
Inference Using Variational Bayes

- **Full conjugacy** allows us to develop a very efficient variational approximation
- Closed-form update equations for all variables
- Proposed method can combine **hundreds or thousands of kernels**

Large-Margin Learning

$$
q(\boldsymbol{f}) = \prod_{i=1}^{N} \mathcal{TN}(f_i; \widetilde{\boldsymbol{e}}^\top \widetilde{\boldsymbol{g}}_i + \widetilde{b}, 1, f_i y_i > \nu)
$$

- $\nu > 0$ corresponds to placing a **margin** between two classes

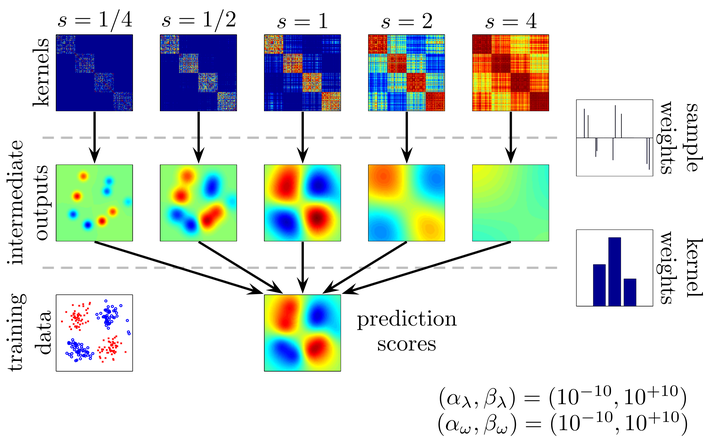


Extensions

- Multiclass learning is done by **sharing kernel weights** in one-versus-all classification
- Semi-supervised learning using **truncated normals** is left for future research

Illustration on a Toy Data Set

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / s^2)$$

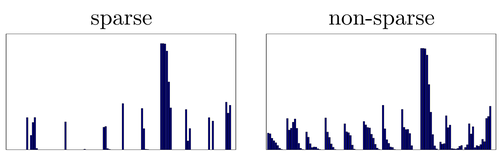


$$(\alpha_\lambda, \beta_\lambda) = (10^{-10}, 10^{+10})$$
$$(\alpha_\omega, \beta_\omega) = (10^{-10}, 10^{+10})$$

Benchmark Data Sets

- 8 benchmark data sets from UCI repository
- Inference takes **less than a minute** with large numbers of kernels, from 91 to 793

	sparse	non-sparse
pima	$N = 537$	$P = 117$
Training Time (sec)	21.15±0.23	20.94±0.22
Test Accuracy (%)	75.02±2.28	74.96±2.08
Selected Kernel (#)	23.20±2.02	79.55±2.93



MKL Data Sets

- 4 comparison data sets for MKL methods
- Protein fold recognition data set

Method	Test Acc.
Damoulas & Girolami (2008)	68.1±1.2
BEMKL (one-versus-all)	**71.5±0.1**
BEMKL (multiclass)	**71.2±0.2**

- Oxford Flowers102 data set

Method	AUC	EER	Acc.
Titsias & Lázaro-Gredilla (2011)	0.952	0.107	40.0
BEMKL (one-versus-all)	**0.969**	**0.068**	**67.0**
BEMKL (multiclass)	**0.969**	**0.069**	**68.9**

Conclusions

- A **Bayesian MKL framework** with a novel kernel combination formulation is introduced
- **Fully conjugate probabilistic model** leads to a very efficient variational approximation
- **Matlab implementation** is available at http://users.ics.aalto.fi/gonen/bemkl

References

Damoulas, T. and Girolami, M. A. Probabilistic multi-class multi-kernel learning: On protein fold recognition and remote homology detection. *Bioinformatics*, 24(10):1264–1270, 2008.

Titsias, M. K. and Lázaro-Gredilla, M. Spike and slab variational inference for multi-task and multiple kernel learning. In *Advances in Neural Information Processing Systems 24*, 2011.

ICML 2012 29th International Conference on Machine Learning, 26 June – 1 July 2012, Edinburgh, Scotland, UK