SUM AND PRODUCT KERNEL REGULARIZATION NETWORKS

Kudová P. and Šámalová T. ({petra,terka}@cs.cas.cz),
Institute of Computer Science, Academy of Sciences of the Czech Republic

Abstract

We study the problem of learning from examples by means of function approximation theory. Based on Aronszajn's formulation of sum of kernels and product of kernels, we derive new approximation schemas – Sum Kernel Regularization Network and Product Kernel Regularization Network. We demonstrate their performance on experiments. For many tasks our schemas outperform the classical solutions.

Motivation

The problem of *learning from examples* is a subject of great interest. It can be formulated as follows. We are given a set of examples $\{(\mathbf{x}_i, y_i) \in R^d \times R\}_{i=1}^N$ that was obtained by random sampling of some real function f, generally in the presence of noise. Our goal is to recover the function f from data, or find the best estimate of it, with respect to generalization.

The problem has been thoroughly studied as a function approximation problem. Since it is ill-posed, regularization techniques are used. We search for a solution as a minimum of the functional composed of error and regularization part:

$$H[f] = \frac{1}{N}\sum_{i=1}^{N}(f(\mathbf{x}_i) - y_i)^2 + \gamma\Phi[f],$$

Φ is a *stabilizer*, $\gamma > 0$ *the regularization parameter.*

We choose a symmetric, positive semi-definite kernel function $K : \Omega \times \Omega \to \mathbb{R}$ (for $\Omega \subseteq \mathbb{R}^d$) and take the corresponding RKHS $\mathcal{H}_K$, with norm $\|\cdot\|_K$. We let the stabilizer be $\Phi(f) = \|f\|_K^2$ and get

$$H[f] = \frac{1}{N}\sum_{i=1}^{N}(f(\mathbf{x_i}) - y_i)^2 + \gamma\|f\|_K^2. \qquad (1)$$

Derivation of the shape of the solution, known as the Representer theorem, has been shown already in [PS03, GJP95, Š04]. All the proofs are based on the idea that a minimum of a function can exist in an interior point only if first derivative equals zero.

Employing the Representer theorem we obtain the solution in the form:

$$f(\mathbf{x}) = \sum_{i=1}^{N} w_i K(\mathbf{x_i}, \mathbf{x}), \qquad (2)$$

where $\mathbf{x}_i$ are the data points and $K(\cdot,\cdot)$ the corresponding kernel. The weights w_i are given by the linear system

$$(N\gamma I + K)\mathbf{w} = \mathbf{y}, \quad \text{where } K_{ij} = K(\mathbf{x_i}, \mathbf{x_j}). \qquad (3)$$

Such function corresponds to a neural network with one hidden layer, called *Regularization Network* (RN)(see Fig. 1).

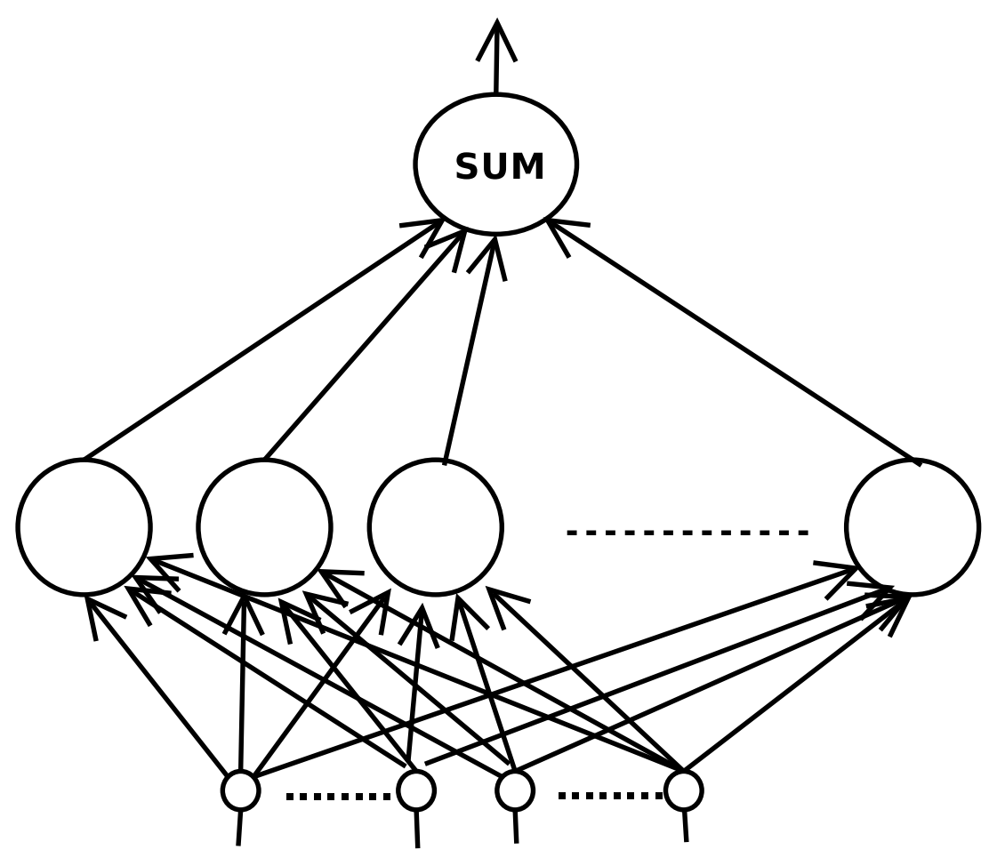

FIGURE 1: Regularization Network schema

Sum and Product Kernels

The choice of the kernel function K is crucial for successful application of the RN. We proposed composite types of kernel functions that may better reflect the data, particularly in cases when the data are heterogenous (have attributes of different types or qualities, differ in different parts of input space).

Theorem 1 *([Aro50]) Let F_i (for $i = 1, 2$) be RKHSs and K_i and $\|.\|_i$ the corresponding kernels and norms. Let $F = \{f \mid f(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x}), f_i \in F_i, f_1 \neq -f_2\}$ and let the norm be $\|f\|^2 = \|\{g'(f), g''(f)\}\|^2 = \|g'(f)\|_1^2 + \|g''(f)\|_2^2$. Then*

$$K(\mathbf{x}, \mathbf{y}) = K_1(\mathbf{x}, \mathbf{y}) + K_2(\mathbf{x}, \mathbf{y}) \qquad (4)$$

is the kernel corresponding to F.
The claim holds also for F defined as class of all functions $f = f_1 + f_2$ with $f_i \in F_i$ and norm $\|f\|^2 = \min(\|f_1\|_1^2 + \|f_2\|_2^2)$ minimum taken for all decompositions $f = f_1 + f_2$ with f_i in F_i.

The kernel function obtained in the theorem 1 we call *Sum Kernel* and corresponding approximation schema *Sum Kernel Regularization Network.*

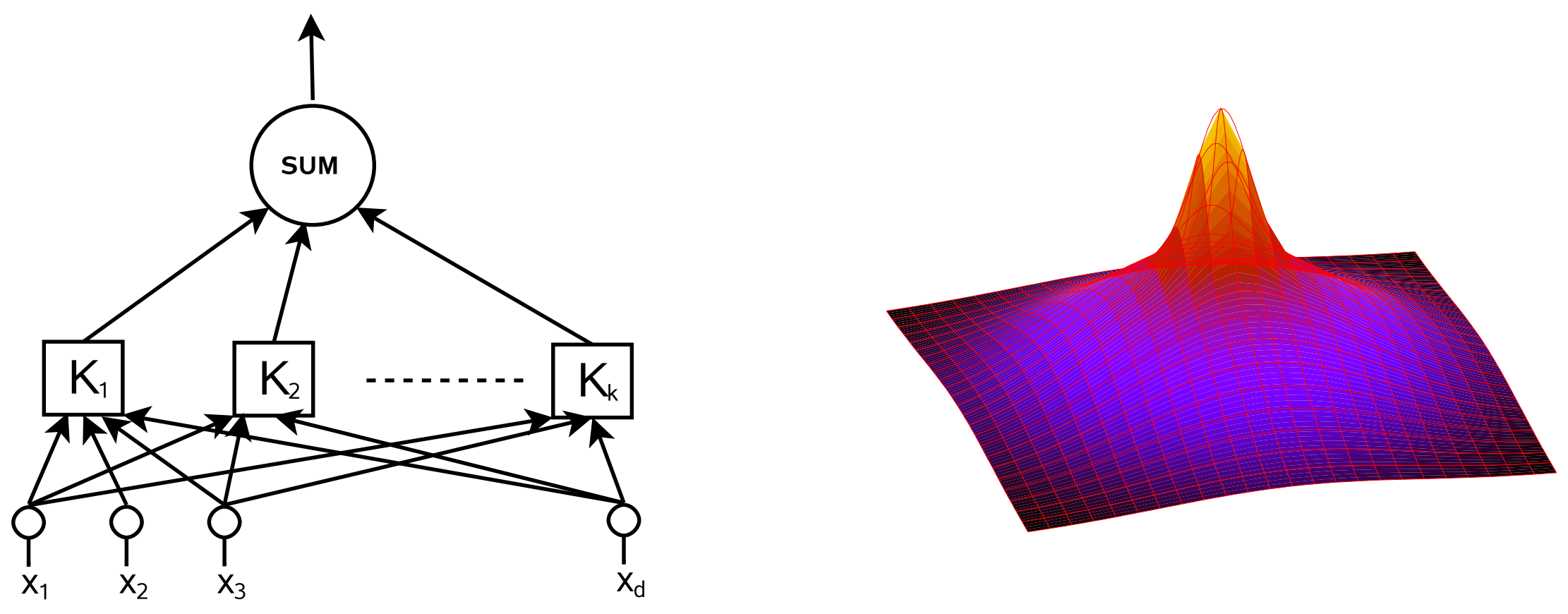

FIGURE 2: An unit realizing sum kernel and an example of such kernel.

Theorem 2 *([Aro50]) For $i = 1, 2$ let F_i be an RKHS on Ω_i with kernel K_i. Then the product $F = F_1 \otimes F_2 = comp\{f \mid f = \sum_{k=1}^{n} f_{1,k}(x_1) f_{2,k}(x_2); n \in \mathbb{N}, f_{i,k} \in F_i\}$ on $\Omega_1 \times \Omega_2$ is an RKHS with kernel given by*

$$K((\mathbf{x_1}, \mathbf{x_2}), (\mathbf{y_1}, \mathbf{y_2})) = K_1(\mathbf{x_1}, \mathbf{y_1})K_2(\mathbf{x_2}, \mathbf{y_2}), \qquad (5)$$

where $\mathbf{x_1}, \mathbf{y_1} \in \Omega_1$, $\mathbf{x_2}, \mathbf{y_2} \in \Omega_2$.

The kernel function from the theorem 2 we call *Product Kernel* and corresponding approximation schema *Product Kernel Regularization Network.*

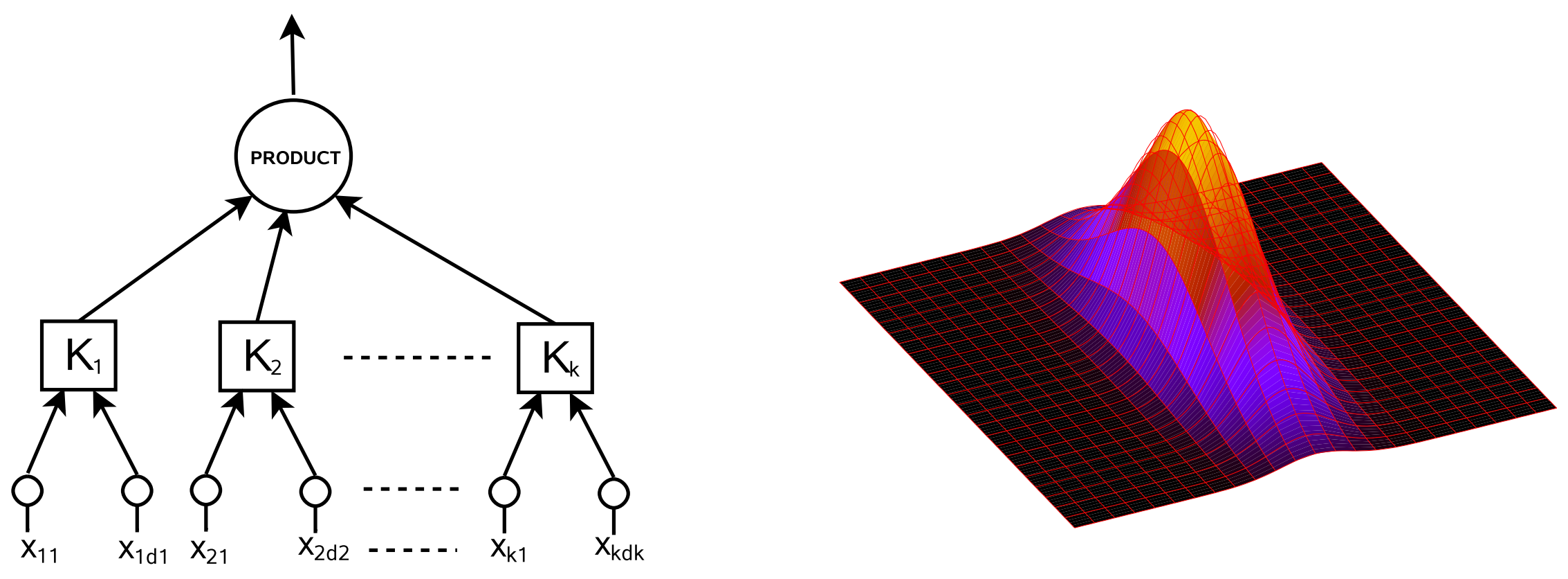

FIGURE 3: An unit realizing a product kernel and an example of such kernel.

Experiments

Described networks were compared on the Proben1 data repository([Pre94]). PKRN was also applied on the prediction of the flow rate on the Czech river Ploučnice.

As Product and Sum Kernels we used products and sums of two Gaussian functions with different widths. The error was always normalized:

$$E = 100\frac{1}{N}\sum_{i=1}^{N}||y_i - f(\mathbf{x_i})||^2.$$

The LAPACK library was used for linear system solving.

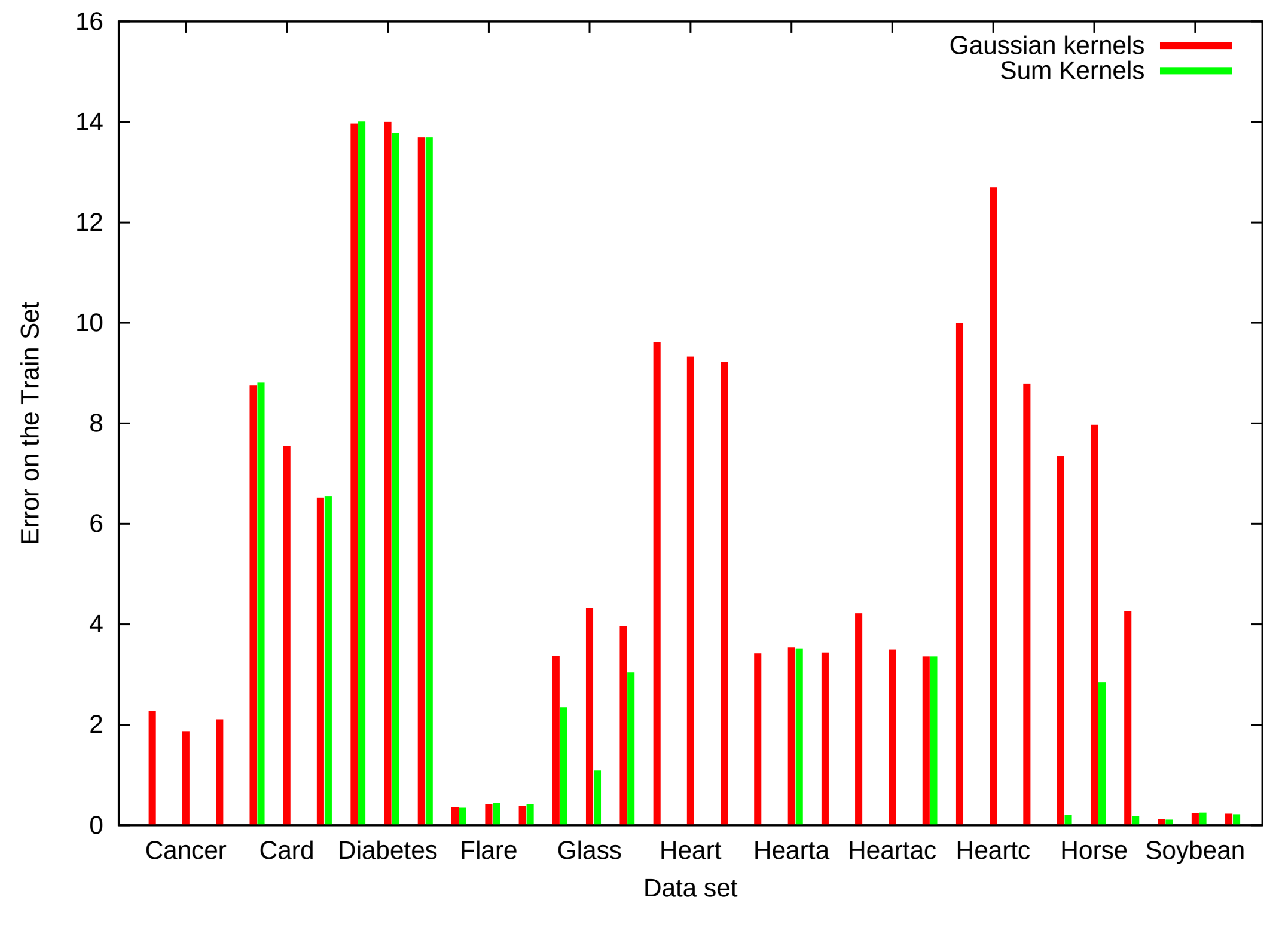

FIGURE 4: Comparison of Sum Kernel and Gaussian kernel: train set error.

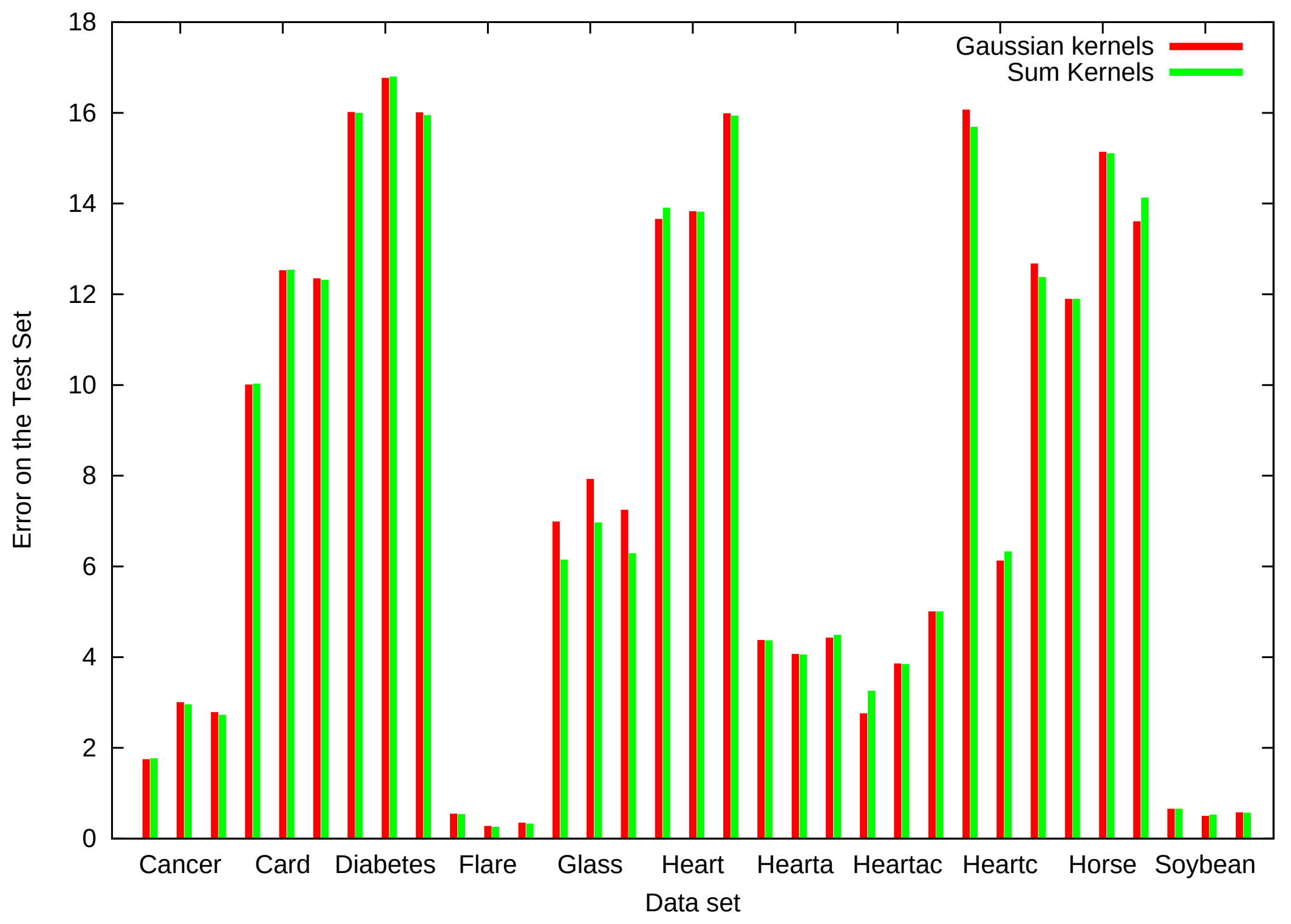

FIGURE 5: Comparison of Sum Kernel and Gaussian kernel: test set error.

	PKRN	CP
E_{train}	0.057	0.093
E_{test}	0.048	0.054

FIGURE 6: Error on the training set E_{train} and error on the testing set E_{test} for prediction of flow rate on the river Ploučnice.

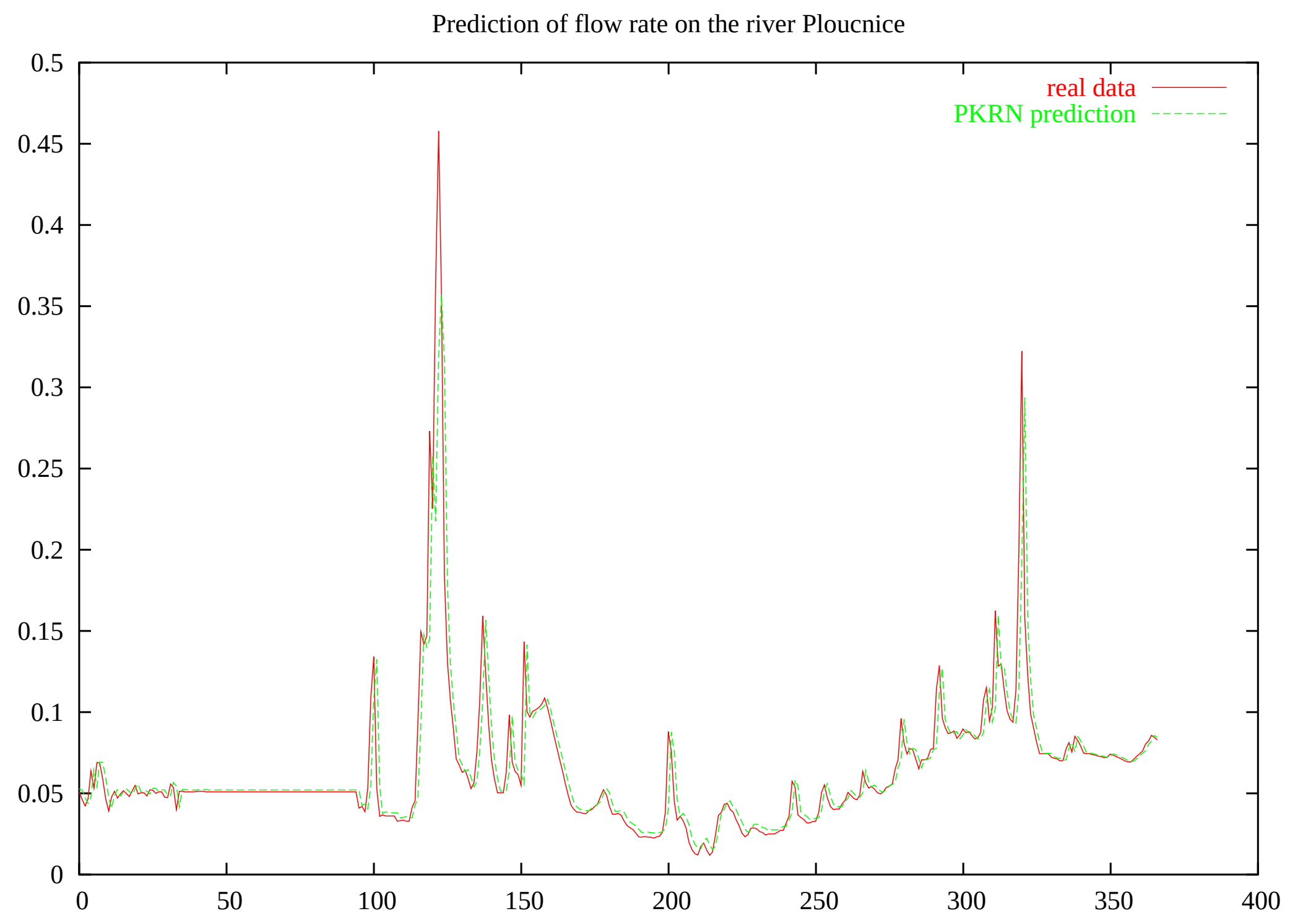

FIGURE 7: Prediction of flow rate by PKRN.

Conclusion

We have shown how to use results of Aronszajn on sums and products of RKHSs to obtain the Sum and Product Regularization Networks.

We compared proposed PKRN and SKRN to classical RN on benchmarks, our SKRN achieved lowest errors in most cases. PKRN was applied on prediction of flow rate on the river Ploučnice and it outperforms the Conservative Predictor.

Our future work should be focused on the application of other types of kernel functions.

References

[Aro50] N. Aronszajn. Theory of reproducing kernels. *Transactions of the AMS*, 68:337–404, 1950.

[GJP95] F. Girosi, M. Jones, and T. Poggio. Regularization theory and Neural Networks architectures. *Neural Computation*, 2:219–269, 7 1995.

[Pre94] L. Prechelt. Proben1 – a set of benchmarks and benchmarking rules for neural network training algorithms. Technical Report 21/94, Universitaet Karlsruhe, 1994.

[PS03] T. Poggio and S. Smale. The mathematics of learning: Dealing with data. *Notices of the AMS*, 50, No.5:537–544, 2003.

[Š04] T. Šidlofová. Existence and uniqueness of minimization problems with fourier based stabilizers. In *Proceedings of Compstat, Prague*, 2004.

P. Kudová is supported by the Program "Information Society" under project 1ET100300414, by the European Commission's Research Infrastructures activity of the Structuring the European Research Area programme, contract number RII3-CT-2003-506079 (HPC-Europa) and also by the Institutional Research Plan AVOZ10300504 "Computer Science for the Information Society: Models, Algorithms, Appplications".
T. Šámalová is supported by the Program "Information Society" under project 1ET100300517 and also by the Institutional Research Plan AVOZ10300504 "Computer Science for the Information Society: Models, Algorithms, Appplications".