Smooth kernel machines for environment and finance

M. Aminian Shahrokhabadi[a], E. Perracchione[b], M. Polato[b], M. Putti[b]

[a]Allameh Tabataba'i University, Teheran; [b]Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova

Acknowledgments: this research has been accomplished within RITA, supported by GNCS-INδAM and UE ERA-PLANET.

Summary

- Machine learning **[Cristianini & Shawe-Taylor, Fasshauer & McCourt]** is a widely used tool for predicting, within a certain tolerance, the evolution of time series, i.e. the dynamics of the considered quantities.

- Focusing on Support Vector Regression (SVR), the main drawback is that measurements are usually affected by noise/errors and might have gaps.

- Since in those cases the learning and prediction steps for capturing the *trend* of time series become very hard, we first construct a reduced kernel-based approximant **[Wirtz et al.]**.

Framework

- Let $\Omega \subseteq \mathbb{R}^d$, $f : \Omega \mapsto \mathbb{R}$ and K be a symmetric and radial kernel. It satisfies the *reproducing property*, namely

$$\langle K(\boldsymbol{x}, \cdot), f \rangle_{\mathcal{H}_K(\Omega)} = f(\boldsymbol{x}),$$

where $\mathcal{H}_K(\Omega)$ denotes the reproducing kernel Hilbert space of K and $\langle \cdot, \cdot \rangle_{\mathcal{H}_K(\Omega)}$ is the inner product.

- Let us introduce the operator $T : L_2(\Omega) \longrightarrow L_2(\Omega)$,

$$T[f](\boldsymbol{x}) = \int_\Omega K(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) d\boldsymbol{y}.$$

By virtue of the Mercer's Theorem we know that the operator T has a countable set of eigenfunctions $\{\varphi_k\}_{k \geq 0}$ (orthonormal in $L_2(\Omega)$) and eigenvalues $\{\lambda_k\}_{k \geq 0}$ so that for $\boldsymbol{x}, \boldsymbol{y} \in \Omega$,

$$K(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k \geq 0} \lambda_k \varphi_k(\boldsymbol{x}) \varphi_k(\boldsymbol{y}), \quad \text{or}$$

$$K(\boldsymbol{x}, \boldsymbol{y}) = \langle \Phi(\boldsymbol{x}), \Phi(\boldsymbol{y}) \rangle_{l_2}, \quad \text{with}$$

$$\Phi(\cdot) = \left(\sqrt{\lambda_1} \varphi_1(\cdot), \sqrt{\lambda_2} \varphi_2(\cdot), \dots \right).$$

- This decomposes K into a *feature* that depends only on $\boldsymbol{x}$ and another one that only depends on $\boldsymbol{y}$. Such expansion is known as the *kernel trick* (see **Fig. 1**).

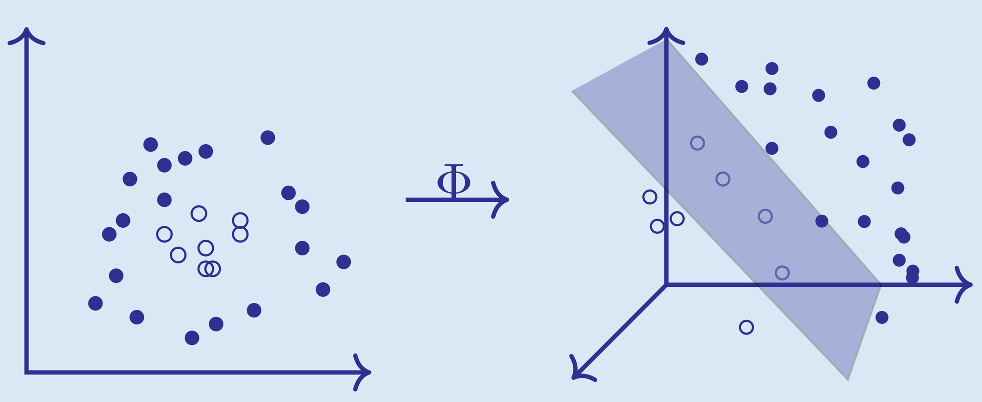

Figure 1. The kernel trick.

- To train the kernel machine in what follows we use a reduced basis method **[Wirtz et al.]**. This leads to a **reduced training** phase.

Method

Given $\mathcal{X}_N = \{\boldsymbol{x}_i, \ i = 1, \dots, N\} \subset \Omega$, a set of distinct data, $\Omega \subseteq \mathbb{R}^d$, with an associated set $\mathcal{F}_N = \{f_i = f(\boldsymbol{x}_i), \ i = 1, \dots, N\}$ of data values, the linear SVR regression model is

$$R(\boldsymbol{x}) = \boldsymbol{x}^\intercal \boldsymbol{w} + b = \sum_{i=1}^N (\alpha_i^* - \alpha_i) \boldsymbol{x}^\intercal \boldsymbol{x}_i + b; \quad \boldsymbol{w}, b, \quad \text{so that} \quad \min_{\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\xi}^*} \left[\frac{1}{2} \boldsymbol{w}^\intercal \boldsymbol{w} + C \sum_{i=1}^N (\xi_i + \xi_i^*) \right],$$

s.t. $R(\boldsymbol{x}_i) - f_i \leq \epsilon + \xi_i$, $f_i - R(\boldsymbol{x}_i) \leq \epsilon + \xi_i^*$, $i = 1, \dots, N$, and $\xi_i \xi_i^* \geq 0$, where C is the *regularization parameter* and ξ_i, $i = 1, \dots, N$, are the slack variables (**Fig. 2**).

Then, for the non-linear regression we use the kernel trick and thus we only need to replace the dot product with the kernel evaluation and the measurements $\boldsymbol{x}_i$ with the feature map $\Phi(\boldsymbol{x}_i)$, $i = 1, \dots, N$. We point out that another drawback of SVR is that it works with fixed length input vectors. Thinking of time series, the model is trained using the very next data point.

Thus, at a certain time step t_k, $k \in \mathbb{N}$, $k > 0$, we are able to predict only t_{k+1}. To address this problem we introduce a sliding window strategy to create the training instances. We refer to this method as Multi SVR (**MSVR**) scheme which is trained via smoother points, constructed only taking a reduced number of bases, namely M.

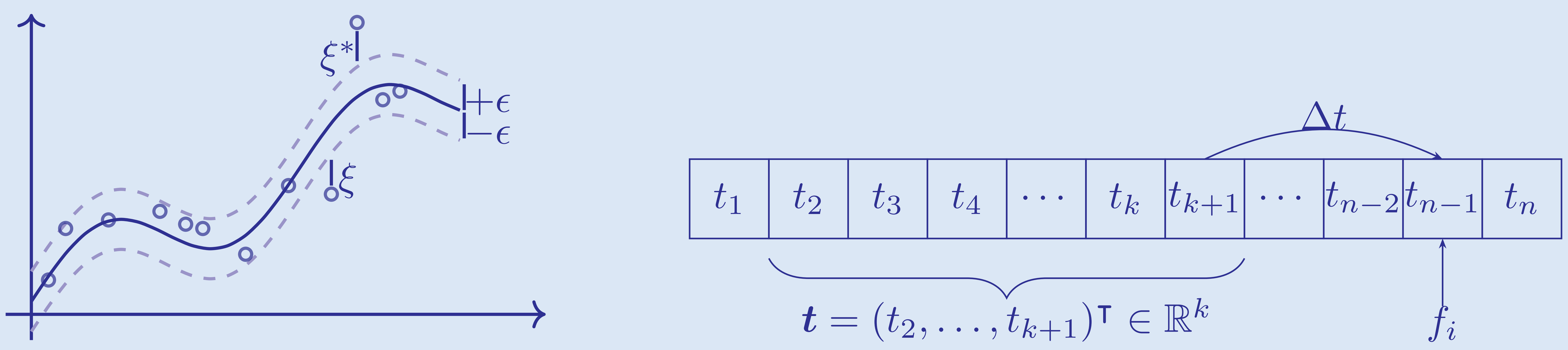

Figure 2. SVR ϵ-tube (left) and MSVR scheme (right).

Results

We plot the considered data sets in **Fig. 3**; available at `http://voss.dmsa.unipd.it/` and `http://www.behranoil.com/`, respectively.

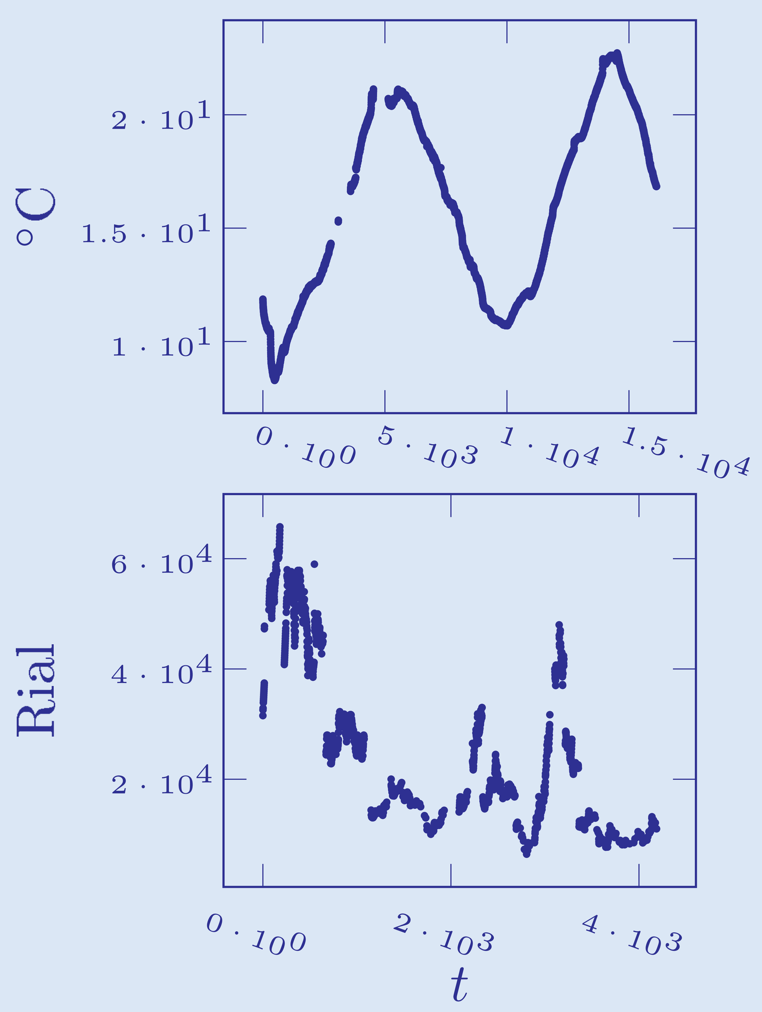

Figure 3. The first data set we consider has been created for an experimental study of the organic soil compaction and prediction of the land subsidence related to climate changes in the South-Eastern area of the Venice Lagoon catchment (**VOSS** - Venice Organic Soil Subsidence). Here we take the temperature in Celsius (°C) sampled one meter below the soil each hour.

The second one belongs to Tehran Securities Exchange Technology Management Co. It reports the volume of daily trades of a stock named Behran Oil, with short name **Shabharn**. Values are reported in *Rial*, the official currency used in Iran.

The Python software is available at `https://github.com/makgyver/vlabtestrepo/`.

					Reduced Training		Classical Training	
Data	N	M	Δt	k	RMSE	RRMSE %	RMSE	RRMSE %
VOSS	14637	393	12	48	0.019	0.034	0.021	0.036
Shabharn	3369	1471	10	30	315.03	0.839	642.05	1.711

Table 1. Results of MSVR approach for environmental and financial data.

References

[1] N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, 2000.
[2] G.E. Fasshauer, M.J. McCourt, *Kernel-based Approximation Methods Using* Matlab, 2015.
[3] D. Wirtz, N. Karajan, B. Haasdonk, *Surrogate modelling of multiscale models using kernel methods*, Int. J. Numer. Met. Eng. **101**, 2015, 1–28.