How to choose the covariance for Gaussian process regression independently of the basis

Workshop *Gaussian Processes in Practice*

Matthias O. Franz and Peter V. Gehler

June 12th, 2006

Motivation: Nonlinear system identification using Volterra series

Characterisation of a nonlinear system $y(t) = T[x(t)]$ by a series expansion $y(t) = \sum_n H_n[x(t)]$ (Volterra, 1887):

$$
\begin{aligned}
y(t) = h^{(0)} &+ \int_{\mathbb{R}} h^{(1)}(\tau_1)x(t-\tau_1)\, d\tau_1 \\
&+ \int_{\mathbb{R}^2} h^{(2)}(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)\, d\tau_1 d\tau_2 \\
&+ \int_{\mathbb{R}^3} h^{(3)}(\tau_1,\tau_2,\tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)\, d\tau_1 d\tau_2 d\tau_2 \\
&+ \cdots
\end{aligned}
$$

Discretised form for $\mathbf{x} = (x_1, \ldots, x_m)^\top \in \mathbb{R}^m$

$$H_n[\mathbf{x}] = \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} h^{(n)}_{i_1 \ldots i_n} x_{i_1} \ldots x_{i_n}.$$

Polynomial regression and Volterra systems

Volterra expansions can be efficiently estimated by a regression in polynomial kernel functions (Franz & Schölkopf, 2006)

$$k_{\mathsf{ihp}}(x, x') = (1 + x^\top x')^p$$

$\Rightarrow$ GP framework is applicable for the estimation of Volterra systems.

Problems:

- Polynomial covariance implies strong correlation of distant inputs. In real-world problems, the reverse situation is more common.
- Typically, polynomial regression shows inferior performance than localized covariance functions.

$\Rightarrow$ Independent choice of covariance and basis

Decoupling of basis and covariance

- Basic idea: approximate a desired covariance function $k_{\mathcal{GP}}(\mathbf{x}_i, \mathbf{x}_j)$ on a finite set $\mathcal{S} = \{x_1, \ldots, x_p\}$ of input points.
- Weight-space view of a GP: $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \Sigma_w \phi(\mathbf{x}_j)$.
- $\Rightarrow$ Choose basis $\phi(\mathbf{x})$ and prior Σ_w such that

$$k_{\mathcal{GP}}(x_i, x_j) = \phi(x_i)^\top \Sigma_w \phi(x_j) \; \forall x_i, x_j \in \mathcal{S}.$$

- Basis: Kernel PCA map $\phi(x) = K^{-\frac{1}{2}}(k(x, x_1), \ldots, k(x, x_n))^\top$, solve system of linear equations in Σ_w.

$\Rightarrow$

- Arbitrary covariances can be approximated.
- Performance of polynomial regression can be significantly improved.