UNIVERSITY OF TECHNOLOGY
IN THE EUROPEAN CAPITAL OF CULTURE
CHEMNITZ

Faculty of Mathematics
Mathematical Finance
Alois Pichler

Regression With Kernels

Perspective: *Seminar thesis, Bachelor, Master's thesis*.

Problem Description

Suppose the function $f\colon \mathbb{R}^d \to \mathbb{R}$ is observed with function values, $(x_i, f_i) \in \mathbb{R}^d \times \mathbb{R}$, $i = 1, \dots, n$. Employing a kernel function $k\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, it holds that

$$\hat{f}(x_i) = f_i, \qquad i = 1, \dots, n, \tag{1}$$

where

$$\hat{f}(x) \coloneqq \sum_{j=1}^{n} k(x, x_j)\, w_j$$

and provided that the weights w satisfy the linear system of equations

$$\sum_{j=1}^{n} k(x_i, x_j)\, w_j = f_i, \qquad i = 1, \dots, n.$$

A canonical candidate for the kernel is the Gaussian kernel $k(x, y) \coloneqq e^{-\|y-x\|^2/\,(2\eta^2)}$ (with Euclidean norm $\|\cdot\|$ in $\mathbb{R}^d$ and characteristic length η).

For observations f_i *with noise*, the *kernel estimator* for $f(x)$ is

$$\hat{f}(x) = \sum_{j=1}^{n} k(x, x_j)\, w_j \tag{2}$$

with

$$\lambda\, w_i + \sum_{j=1}^{n} k(x_i, x_j)\, w_j = f_i, \qquad i = 1, \dots, n. \tag{3}$$

The regression parameter λ is chosen to match the noise: $\lambda = 0$ matches the initial situation so that $\hat{f}(x_i) = f_i$ without noise (interpolation, cf. (1)). A value of λ too close to zero leads to *overfitting*, larger values for λ lead to *oversmoothing* (underfitting). λ acts as a regularization parameter, balancing bias and variance to prevent overfitting while maintaining model flexibility.

Task: Delayed Embedding of Dimension d

We shall employ (2) and (3) to a stationary time series $X=(X_t)$ with $X_t\in\mathbb{R}$, $\mathbb{E}\,X_t=0$, and constant variance

$$\sigma^2 \coloneqq \operatorname{var}\bigl(X_{t+1}|\, X_{t-d+1},\dots,X_t\bigr),$$

where $d\in\mathbb{N}$ is fixed.

Consider the observation (realization)

$$X_{2-d},\dots,X_1,X_2,\dots,X_{n+1} \tag{4}$$

of the time series. In the *sliding window approach*, set

$$x_i \coloneqq (X_{i-d+1},X_{i-d+2},\dots,X_i)\in\mathbb{R}^d \quad \text{and} \quad f_i \coloneqq X_{i+1}\in\mathbb{R}, \qquad i=1,\dots,n, \tag{5}$$

then $\hat f\bigl((X_{t-d+1},\dots,X_t)\bigr)$ is an *estimate* for the subsequent X_{t+1},

$$\hat f\bigl((X_{t-d+1},\dots,X_t)\bigr)\approx \mathbb{E}\bigl[X_{t+1}|\, X_{t-d+1},\dots,X_t\bigr].$$

1. Discuss and implement the relations (2), (3) and (5) for some adequate, oscillating time series observations (X_t) as in (4).

2. **Recursive multistep forecasting:** For some chosen starting values $(\hat X_1,\dots,\hat X_d)$, extend this time series by setting

   $$\hat X_{t+1}\coloneqq \hat f\bigl((\hat X_{t-d+1},\dots,\hat X_t)\bigr)+\varepsilon_t, \qquad t=d,\, d+1,\dots$$

   and *plot* the realization $\hat X_1,\hat X_2,\dots,\hat X_d,\dots,\hat X_n,\dots,\hat X_N$ with some $N\gg d$; the iid random errors ε_t have 0-mean, and variance corresponding to the residual variance,

   $$\operatorname{var}\varepsilon_t\approx\sigma^2=\operatorname{var}\bigl(X_{t+1}|\,(X_{t-d+1},\dots,X_t)\bigr).$$

3. Adjust the parameters λ, d, σ and the kernel k (or the parameter in the kernel, η, say) so that the time series $\hat X$ *visually* matches the initial observations (4) of X.

4. Try the norm $\|x\|_\Sigma\coloneqq x^\top\Sigma^{-1}x$ (Mahalanobis distance) instead of the Euclidean norm, where Σ is the covariance matrix of the random vector $(X_{i-d+1},\dots,X_i)$. The covariance matrix can be estimated from the observations (5).

Bonne chance!