Gaussian processes and kernel smoothing methods.

Datasets

- HydroGeology.

Problems

Select any function from the table below. Generate a sample $(x_1, \ldots, x_N)$ from the uniform distribution on the interval defined by $\mathrm{dom}(f)$. Next, generate $y_n = f(x_n) + \varepsilon_n$, where $\varepsilon_n \sim \mathcal{N}(0, \sigma^2), n = 1, \ldots, N$. We assume that y is a dependent variable, and x is a predictor. Then do the following:

1. Estimate f using kernel smoothing method. Try to minimize the cross-validation or similar score by choosing an appropriate kernel, bandwidth and local fit type.

2. Estimate f using Gaussian process regression. Try to minimize the cross-validation or similar score by choosing an appropriate kernel and its parameters.

For the HydroGeology dataset, use the Gaussian process and kernel smoothers to predict the value of variable q. Try to choose the parameters so as to minimize the cross validation error or similar score.

Recommendations

Kernel smoothing can be applied using `ksmooth` function of `stats` package or `locpoly` function of `KernSmooth` package. More sophisticated functional, which also supports multidimensional smoothing, can be found in `np` package. The functions of interest are `npregbw` (calculates optimal bandwidth) and `npreg` (actual regression implementation). Calling `vingette("np")` before using this package can be most helpful.

Gaussian process regression and classification can be done using `gausspr` function of `kernlab` package.

$f(x)$	$\mathrm{dom}(f)$	σ	N
$\frac{\sin(x)}{x}$	$[0; 4\pi]$	0.2	50
$\cos x + \frac{x}{4}$	$[0; 4\pi]$	0.75	50
$\exp(-x^2) - \exp(-\frac{(x-2)^2}{8})$	$[-2; 4]$	0.2	50
$\exp(-x) + \frac{1}{4}\sin(4x)$	$[0; 4]$	0.1	50
$x \cos x$	$[0; 2\pi]$	1	50
$\ln\left(\frac{x}{\pi - x}\right) - \sin(4x)$	$(0; \pi)$	0.75	50
$\ln\left(\sin(x^2) + 1.2\right)$	$[0; 4\pi]$	0.25	50
$x^2 + \sin(2\pi x) I(x > 0)$	$[-2; 2]$	0.5	50
$\sqrt{\lvert x \rvert} - x\cos(\pi x) I(x > 0)$	$[-2; 2]$	0.4	50
$\ln(x)\sin(\pi x)$	$[0; 2]$	0.15	50
$\sin(x)\ln(1 + x^2)$	$[-\pi; \pi]$	0.5	50