58 Modelling biological systems with nonparametric models (1)

Project level: Moderate/Difficult

58.1 Overall project description

Biological systems, such as prey-predator models, often exhibit complex non-linear relationships that parametric statistical models struggle to describe. Nonparametric models are more suited to modelling such systems as their structure provides a more flexible modelling framework.

In this project, you will investigate and compare nonparametric statistical modelling methods on datasets obtained from biological dynamic systems. In particular, you will learn and fit the smoothing splines model and compare it with some of the more well-known nonparametric model such as the Reproducing Kernel Hilbert Space (RKHS) regression model.

This project will provide you with excellent training experience in advanced machine learning algorithms. R packages (KGode) are available to carry out the simulation study, but there is scope for more experienced programmers to code the simulations yourself.

58.2 Individual project details

How many individual projects are available in this area: 1.

Data available

Data can be generated from the R package KGode by running the code below. The ODE (ordinary differential equation) model used here is the Lotka-Volterra model (prey-predator). From the simulated dataset below, we have y_no, the noised observation of the two states variables in the ODE model. And t_no, the time points for each observation.

```
require(mvtnorm)
library(KGode)

noise = 0.1  ## set the variance of noise
SEED = 19537
set.seed(SEED)

## Define ode function, we use lotka-volterra model in this example.
## we have two ode states x[1], x[2] and four ode parameters alpha, beta,
## gamma and delta.
LV_fun = function(t,x,par_ode){
  alpha=par_ode[1]
```

```
  beta=par_ode[2]
  gamma=par_ode[3]
  delta=par_ode[4]
  as.matrix( c( alpha*x[1]-beta*x[2]*x[1] , -gamma*x[2]+delta*x[1]*x[2] ) )
}

## create a ode class object
kkk0 = ode$new(2,fun=LV_fun)
```

```
## ode is sample 2.
```

```
## set the initial values for each state at time zero.
xinit = as.matrix(c(0.5,1))
## set the time interval for the ode numerical solver.
tinterv = c(0,6)
## solve the ode numerically using alpha=1, beta=1, gamma=4, delta=1.
kkk0$solve_ode(c(1,1,4,1),xinit,tinterv)
## Add noise to the numerical solution of the ode model and
## treat it as the noisy observation.
y_true= t(kkk0$y_ode)
n_o = max( dim( y_true) )
t_no = kkk0$t
y_no = y_true + rmvnorm(n_o,c(0,0),noise*diag(2))
```

Students are also encrouaged to try other ODE models such as the SIR model which is widely used in epidemiology.

Question(s) of interest
The main questions of interest are:

- How does the smooth spline model perform in comparison with the more commonly used parametric models such as linear regression models?
- How does the smooth spline model perform in comparison with the other nonparametric models such as RKHS?

Relevant courses
We strongly recommend that you have taken the following courses to undertake this project:

- Linear regression models.
- Flexible regression.