CC Research - Process (log)

Gulzina Kuttubekova

Fall 2018 - Spring 2019

Assumptions

1. We want to develop a model which will serve as an emulator. To test the accuracy of an emulator and modify/optimize the model later, we need to simulate data. Utilizing Gaussian processes, we simulated data from GP in the following manner:

   (a) Simulate $\underset{\sim}{x}_i \sim GP(0, \Sigma)$ for $i = 1, ..., n$, n is the sample size and $\underset{\sim}{x}_i$ is a vector of length $T + 1$, i.e $\underset{\sim}{x}_i = (x_0, x_1, ..., x_T)$, where Σ is AR(1) with unit variance ($\sigma = 1$) s.t $\Sigma = \begin{bmatrix} 1 & \rho & \rho^2 & & \\ \rho & 1 & \rho & . & . \\ \rho^2 & \rho & 1 & & \\ & . & & . & \\ & . & & & . \end{bmatrix}$

   ρ is fixed ro be 0.95.

   Note: We assume that Time(T) $<<$ sample size(n), where both T and n are integers.

   (b) **Initially**, we defined weight parameter $w(t)$ to be deterministic s.t

   $$w(t) = \frac{T + 1 - t}{T + 1}$$

where $T \geq 1$ and $0 \leq t \leq T$. Estimates of $w(t)$ which were found by using this assumption, were pretty close to the true ones. However, taking into account correlated structure of weights, we introduced new structure for weights:

Later, we defined weight parameter to be sampled from GP with zero mean and AR(1) covariance matrix s.t

$$w_t \sim GP(0, \Pi)$$

where Π has AR(1) structure.

More about the distribution of w:

$$p(w) = N(w; 0, \Pi)$$

where $\Pi = \frac{1}{1-\psi^2}\Sigma$, and $\Sigma = \begin{bmatrix} 1 & \psi & \psi^2 & & \\ \psi & 1 & \psi & . & . \\ \psi^2 & \psi & 1 & & \\ & . & & . & \\ & . & & & . \end{bmatrix}$ We need to find the value of ψ and σ^2 s.t for $|\rho| < 1$

$$w_t = \rho w_{t-1} + z_t$$

$$z_t \overset{indep}{\sim} N(0, \sigma^2)$$

Since we want a stationary process, i.e time independent covariance between w_t and w_{t+h}, we perform the following steps to find the appropriate covariance matrix:

$$w_0 \sim N(0, \widetilde{\sigma}^2)$$

$$w_1 = \rho w_0 + z_1$$

which implies

$$w_1 \sim N(0, \rho^2 \widetilde{\sigma}^2 + \sigma^2)$$

where we want $\widetilde{\sigma}^2 = \rho^2 \widetilde{\sigma}^2 + \sigma^2$ s.t $\widetilde{\sigma}^2 = \frac{\sigma^2}{1-\rho^2}$. Hence,

$$(w_0, w_1) \sim N\Big(0, \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \Big)$$

In the same manner, we found the entire distribution of

$$(w_0, w_1, ..., w_T) \sim N\Bigg(0, \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & & \\ \rho & 1 & \rho & . & . \\ \rho^2 & \rho & 1 & & \\ & . & & . & \\ & . & & & . \end{bmatrix} \Bigg)$$

For convenience, we set the white noise variance $\boldsymbol{\sigma^2 = 1}$ and increased the penalty $\boldsymbol{\rho = 0.99}$

(c) As a next step we calculate a $n \times n$ covariance matrix Ω: correlations between x_i and x_j $\forall i \neq j$, according to the assumptions we provided earlier, i.e

$$\Omega_{ij} = e^{-d(x_i,x_j)\theta}$$

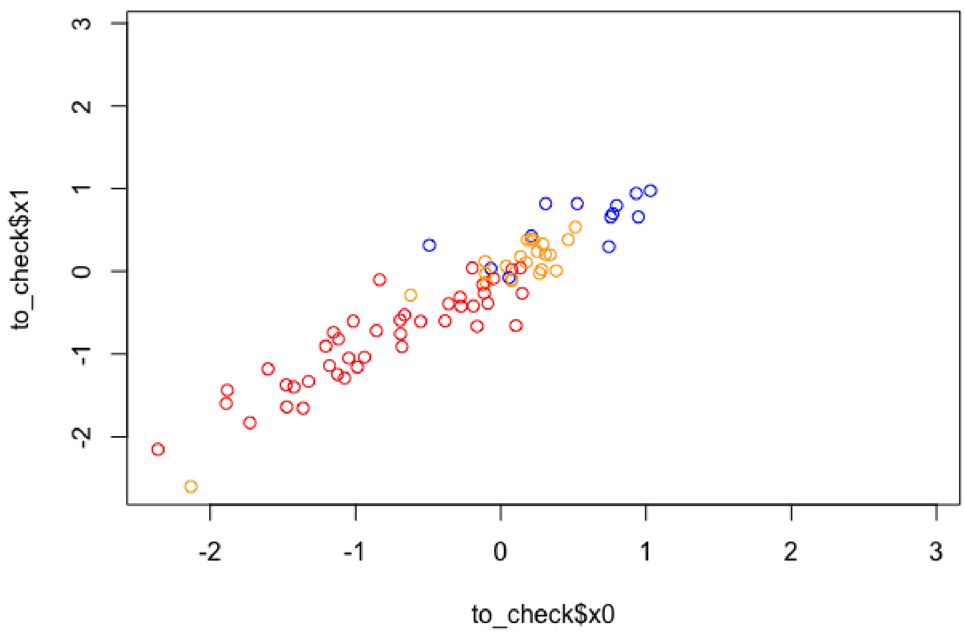
where $\theta = 1$ and

$$d(x_i, x_j) = \sum_{t=0}^{T} w_t (x_{it} - x_{jt})^2$$

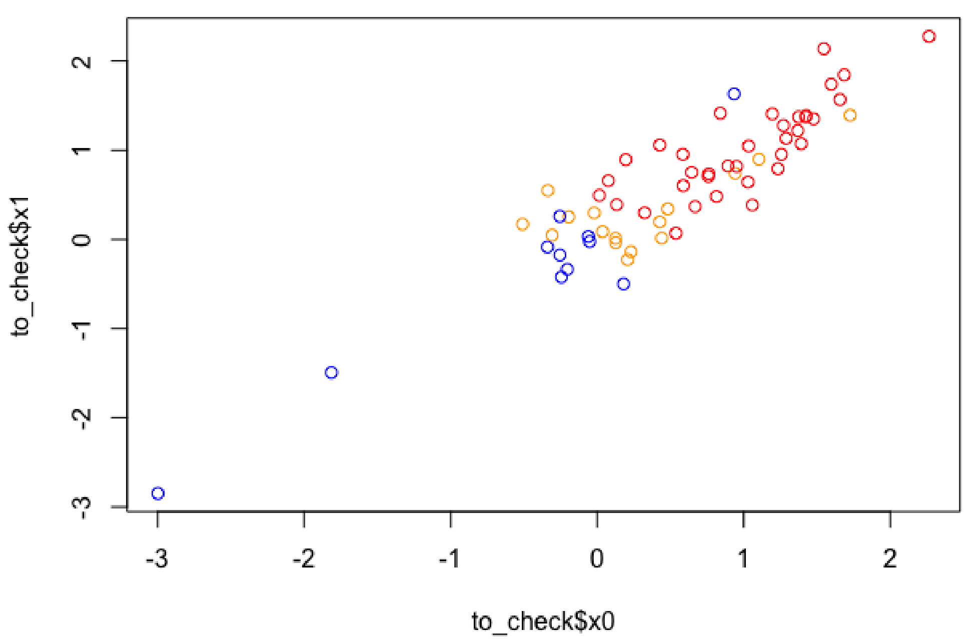
(d) Finally, sample of size n of response variables y, here y is assumed to be scalar, are simulated from another Gaussian Process:

$$y_i \sim GP(0, \Omega)$$

2. As a next step, we want to know whether the simulated sample of responses are fair enough (reasonable) and truly come from the model that we have assumed previously. For that reason, we have developed series of "checkers". Using those "checkers" we can visually analyze the simulated data and tell whether it is reasonable. Sample code is given in the *test_on_simulation.R* file. Also some results have been attached:



to_check$x1
to_check$x0



to_check$x1
to_check$x0