

Gaussian Process regression (Kriging)

Georgios Karagiannis

Department of Mathematics, Purdue

June 28, 2016

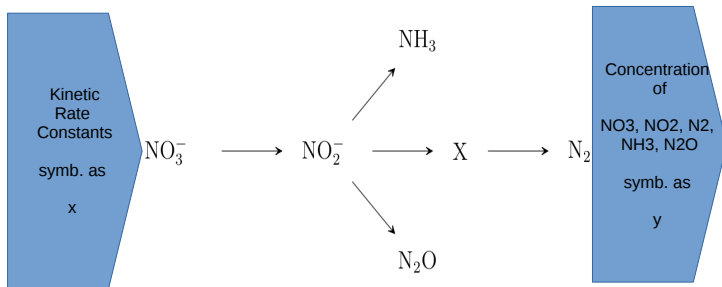
SURF 2016

Why is it useful in UQ ?

E.g.: consider the framework of Computer Experiments:

- There is an expensive Simulator that describes a Physical procedure

Catalytic Conversion of Nitrate to Nitrogen



Why is it useful in UQ ?

It can be used as an emulator (probabilistic surrogate) in:

- Prediction of the output of expensive simulators (***)
- Optimization of expensive objective functions
- Sensitivity Analysis in expensive simulators
- Uncertainty Propagation in expensive simulators

Preliminaries

Multivariate Normal distribution (I)

Notation

$$f \sim N_n(\mu, \Sigma), \quad f := (f_1, \dots, f_n)^\top$$

Mean (vector)

$$\mu := (E(f_1), \dots, E(f_n))^\top, \quad \mu_i = E(f_i)$$

Covariance (matrix)

$$\Sigma = \begin{bmatrix} \text{Cov}(f_1, f_1) & \cdots & \text{Cov}(f_1, f_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(f_n, f_1) & \cdots & \text{Cov}(f_n, f_n) \end{bmatrix}, \quad \Sigma_{i,i'} = \text{Cov}(f_i, f_{i'})$$

Multivariate Normal distribution (II)

Notation

$$f \sim N_n(\mu, \Sigma), \quad f := (f_1, \dots, f_n)^\top$$

Density function

$$N_n(f|\mu, \Sigma) = \frac{1}{\sqrt{2\pi|\Sigma|}} \exp\left(-\frac{1}{2}(f - \mu)^\top \Sigma^{-1}(f - \mu)\right)$$

Cumulative function

$$\Pr(f \leq U|\mu, \Sigma) = \int_{-\infty}^U N_n(f|\mu, \Sigma) df$$

Multivariate Normal distribution (III)

If

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \sim N_{n_1+n_2} \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12}^T \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix} \right)$$

then conditioning implies

$$f_1 | (f_2 = t) \sim N_{n_1}(\mu_{1|2}, \Sigma_{1|2})$$

where

$$\mu_{1|2} = \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(t - \mu_2)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T$$

Gaussian process (GP)

Definition GP is a collection of random variables $\{f(x); x \in \mathcal{X}\}$, indexed by label x , where any finite collection of those variables has a multivariate normal distribution

Namely We denote the GP as

$$f(\cdot) \sim \text{GP}(\mu(\cdot), c(\cdot, \cdot))$$

with mean

$$\mu(x) := \mathbb{E}(f(x))$$

and covariance function

$$c(x, x') := \text{Cov}(f(x), f(x'))$$

Notes Essentially, GP is a distribution defined over functions
GP is specified by its mean and covariance functions.

Gaussian process regression (Krigin)

The prior GP model

Prior information about $f(\cdot)$ is represented as a GP, as:

$$f(\cdot) | \beta, r, \tau^2, \sigma^2 \sim \text{GP}(\mu_0(\cdot), c_0(\cdot, \cdot)),$$

with

- mean function

$$\mu_0(x) = \sum_{k=0}^p \beta_k h_k(x) = h^T(x) \beta$$

$$\text{E.g., } \mu_0(x) = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \dots + \beta_p x^p$$

- covariance function

$$c_0(x, x') := \tau^2 \prod_{j=1}^d R_j(x_j, x'_j | r_j) + \delta_{x, x'} g$$

Some $R(x, x'|\psi)$ leading to valid covariance functions

- Gaussian covariance functions:

$$R(x, x'|r) = \exp\left(-\frac{1}{2} \frac{|x - x'|^2}{r^2}\right)$$

- Exponential covariance function

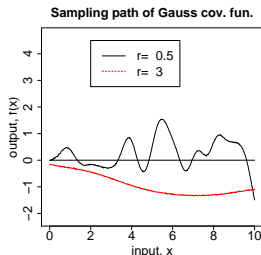
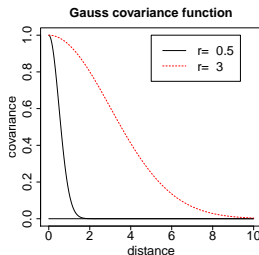
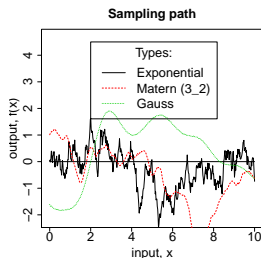
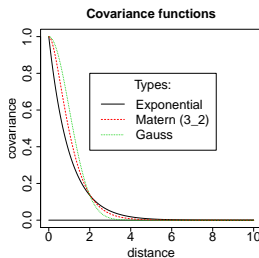
$$R(x, x'|r) = \exp\left(-\frac{|x - x'|}{r}\right)$$

- Matern covariance function

$$R(x, x'|r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|x - x'|}{r}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}|x - x'|}{r}\right)$$

K_ν modified Sobol function, for $\nu = 3/2, 5/2, \dots$

Running example



The statistical model (likelihood function)

- Suppose available training data-set $D = \{(x_i, y_i); i = 1, \dots, n\}$

$$y_i = \underbrace{f(x_i)}_{f(\cdot) \sim \text{GP}(\mu_0(\cdot), c_0(\cdot, \cdot))} + \underbrace{\epsilon_i}_{\epsilon_i \sim \text{N}(0, \sigma^2)}$$

The statistical model (likelihood function)

- Suppose available training data-set $D = \{(x_i, y_i); i = 1, \dots, n\}$

$$y_i = \underbrace{f(x_i)}_{f(\cdot) \sim \text{GP}(\mu_0(\cdot), c_0(\cdot, \cdot))} + \underbrace{\epsilon_i}_{\epsilon_i \sim \text{N}(0, \sigma^2)}$$

- The joint distribution of \mathbf{y} (a.k.a. likelihood function) is

$$\mathcal{L}(\mathbf{y} | \beta, r, \tau^2, \sigma^2) = \text{N}_n(\mathbf{y} | \boldsymbol{\mu}, \mathbf{C} + \mathbb{I}_n \sigma^2)$$

$$\mathbf{y} = [y_1 \quad \cdots \quad y_n]^\top;$$

$$\boldsymbol{\mu} = [\mu_0(x_1) \quad \cdots \quad \mu_0(x_n)]^\top;$$

$$\mathbf{C} = \begin{bmatrix} c_0(x_1, x_1) & \cdots & c_0(x_1, x_n) \\ \vdots & \ddots & \vdots \\ c_0(x_1, x_n) & \cdots & c_0(x_n, x_n) \end{bmatrix}$$

Towards a probabilistic surrogate model

- Let $f(x)$ & $f(x')$ be 'future' function values at inputs x & $x' \in \mathcal{X}$.
- Then joint distribution of $(f(x), f(x'), y)^\top$ is

$$\begin{bmatrix} f(x) \\ f(x') \\ \mathbf{y} \end{bmatrix} \sim N_{n+2} \left(\begin{bmatrix} \mu_0(x) \\ \mu_0(x') \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} c_0(x, x) & c_0(x, x') & \mathbf{c}(x)^\top \\ c_0(x', x) & c_0(x', x') & \mathbf{c}(x')^\top \\ \mathbf{c}(x) & \mathbf{c}(x') & \mathbf{C} + \mathbb{I}\sigma^2 \end{bmatrix} \right)$$

where $\mathbf{c}(x) = (c_0(x, x_i); i = 1, \dots, n)^\top$

The Posterior GP model (probabilistic surrogate model)

By conditioning on \mathbf{y} , it can be shown that $f(\cdot)|D, \beta, L, \tau^2, \sigma^2$ is a GP

$$f(\cdot)|D, \beta, r, \tau^2, \sigma^2 \sim \text{GP}(\mu_n(\cdot), c_n(\cdot, \cdot))$$

with ...

The Posterior GP model (probabilistic surrogate model)

By conditioning on \mathbf{y} , it can be shown that $f(\cdot)|D, \beta, L, \tau^2, \sigma^2$ is a GP

$$f(\cdot)|D, \beta, r, \tau^2, \sigma^2 \sim \text{GP}(\mu_n(\cdot), c_n(\cdot, \cdot))$$

with ...

- mean function

$$\mu_n(x) = \mu_0(x) + \mathbf{c}(x)(\mathbf{C} + \mathbb{I}\sigma^2)^{-1}(\mathbf{y} - \boldsymbol{\mu})$$

- covariance function

$$c_n(x, x') = c_0(x, x') + \mathbf{c}(x)(\mathbf{C} + \mathbb{I}\sigma^2)^{-1}\mathbf{c}(x')^\top$$

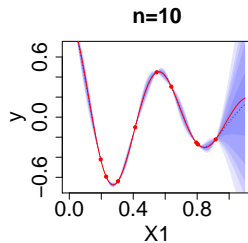
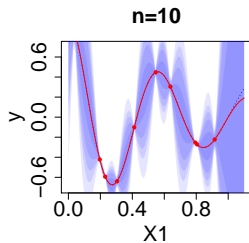
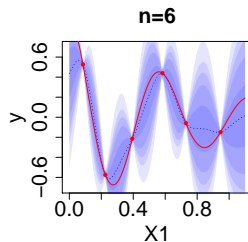
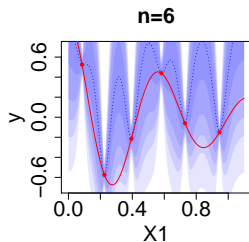
Recall:

$$\boldsymbol{\mu} = (\mu_0(x_1), \dots, \mu_0(x_n))^\top,$$

$$\mathbf{C} = (c_0(x_i, x_j); i = 1, \dots, n, j = 1, \dots, n)$$

$$\mathbf{c}(x) = (c_0(x, x_1), \dots, c_0(x, x_n))^\top$$

Running example



How to train the GP regression ?

How to learn $\beta, r, \tau^2, \sigma^2$?

- Classical statistical inference:

E.g by Maximum Likelihood Estimation (MLE):

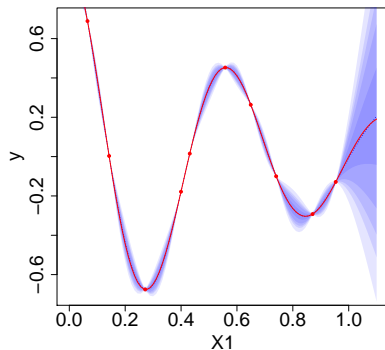
$$(\hat{\beta}, \hat{r}, \hat{\tau}^2, \hat{\sigma}^2) = \arg \min_{\forall(\beta, \ell, \tau^2, \sigma^2)} (-2 \log(\mathcal{L}(\mathbf{y}|\beta, r, \tau^2, \sigma^2)))$$

- Bayesian statistical inference:
 - By Maximum A posteriori Estimation (MAP)
 - By evaluating the posterior distributions

Here, we focus on MLE ... easier to digest

Running example

Posterior (trained) GP regression



Mean $\mu_n(\cdot)$:

Estimate (Intercept) 0.3460

Covariance $c_n(\cdot, \cdot)$:

Type : Matern $\nu = 5/2$

Estimate \hat{r} : 0.2846

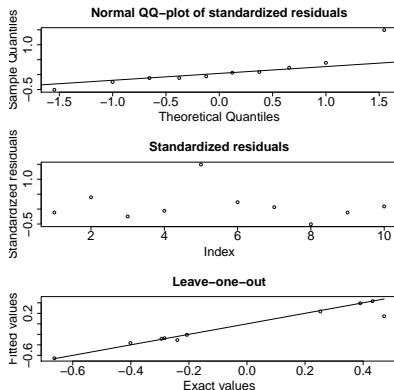
Variance $\hat{\sigma}^2$: 0.6674608

Nugget effect \hat{g} : 1e-07

How to assess the GPR model ?

Check for:

- Normality assumption
- Goodness of fit
- Predictive ability



Compare different models (Eg. Gauss vs Matern cov. funct.)

Leave-one-out Cross Validation (LOO-CV)

For $j = 1, \dots, n$:

- 1 Train the GP regression model against data-set $D^{(-j)} = \{(x_i, y_i); \forall i \neq j\}$
- 2 Predict \hat{y}_j at input x_j , based on the GP regression model
- 3 Compute a performance criterion ($C_j = C(\hat{y}_j, y_j)$) measuring how close your predictions (\hat{y}_j) to the real values (y_j) are.
 - E.g.: R^2 , MSE, MAE

Compute the cross validation criterion, as $CV = \frac{1}{n} \sum_{i=1}^n C_i$

Performance criteria

- Coefficient of determination :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- The percentage of the total variation explained by the predictions

- Root mean squared error :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|^2}$$

- penalizes larger differences

- Mean absolute error:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- more robust to outliers

Running example

| | | LOO-CV criterion | | |
|-----|-------------|------------------|------|------|
| | | R2 | RMSE | MAE |
| GPR | Model | | | |
| | Gaussian | 0.99 | 0.03 | 0.01 |
| | Exponential | 0.45 | 0.29 | 0.21 |
| | Matern 5/2 | 0.96 | 0.07 | 0.04 |

Table: Model comparison

And now,

let's practice ...