Gaussian process model selection for computer experiments
Sébastien Petit, Julien Bect, Paul Feliot, Emmanuel Vazquez

To cite this version:
Sébastien Petit, Julien Bect, Paul Feliot, Emmanuel Vazquez. Gaussian process model selection for computer experiments. MASCOT PhD student 2020 Meeting, Sep 2020, Grenoble, France. hal-03018559

HAL Id: hal-03018559
https://centralesupelec.hal.science/hal-03018559
Submitted on 22 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives| 4.0 International License
Gaussian process model selection for computer experiments

Sébastien J. Petit, Julien Bect, Paul Feliot, and Emmanuel Vazquez

1Université Paris-Saclay, CentraleSupélec, Laboratoire des Signaux et Systèmes, Gif-sur-Yvette, France.
2Safran Aircraft Engines, Moissy-Cramayel, France.

email: sebastien.petit@centralesupelec.fr

Context
- Exploration of black-box numerical simulations $f: \mathbb{R}^d \rightarrow \mathbb{R}$ with Gaussian processes
- Given data $D_n = (\mathbf{x}_i, f(\mathbf{x}_i))$, a Gaussian process $\xi$ can be used to make probabilistic predictions of $f$
- $\xi$ is a prior over functions
- The choice of $\xi$ is critical for good predictions and design-of-experiments techniques

The prior $\xi$ is often chosen within a parametric family.
- Often, the Matérn covariance function family is used
- Many procedures have been proposed in the literature for selecting the parameters of a covariance function
- Little is known about their relative benefits

What are the most useful procedures to select the parameters of a Matérn covariance function (including or not regularity)?

1 Maximum-likelihood [5]
- A very popular technique
- Choose the parameters that yield the highest value of the probability density for the observations, or equivalently, minimize

$$ L(\beta, \phi) = - \frac{1}{2} \sum_{i=1}^{n} \ln(2\pi) + \ln(\det(K_0)), $$

where $K_0$ is the covariance matrix of $\mathbf{z}$ at points $\mathbf{x}_i = (x_1, \ldots, x_d)$ for parameters $\beta$ and $\phi = (\phi_1, \ldots, \phi_d)$ denotes the value of $f$ at $\mathbf{x}_i$.

2 Cross-validation
- Leave-one-out (LOO) [6] is a second very popular technique
- Consists in averaging losses for predicting one observation using the others.
- We suggest using negatively-oriented scoring rules [4] for the loss functions
- A negatively-oriented scoring rule is a mapping $\mathcal{S}(\mathbb{P}, P) \rightarrow \mathbb{R}$ where $\mathbb{P}$ is a class of probability distributions and $P(\mathbb{P})$ represents a loss for observing $z$ while predicting $P$
- Given a scoring rule $\mathcal{S}$ the corresponding LOO criterion is

$$ L(\beta, \phi) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{S}(P(z_i), P(\mathbf{x}_i)), $$

where $\mathcal{S}(P(z_i), P(\mathbf{x}_i))$ denotes LOO predictive distributions.

In this we consider the following scoring rules [4]
- $\mathcal{S}(P(z), Y) = (\mathbb{E}_{P(z)}(Y) - Y)^2$
- $\mathcal{S}(P(z), Y) = \mathbb{H}(z) - \mathbb{H}(Y)$, with $\mathbb{H}(z)$ the pdf of $P$
- $\mathcal{S}(P(z), Y) = \mathbb{H}(Y) - \mathbb{H}(z)$, with $P$ the cdf of $P$

We shall denote the resulting selection procedures by LOO-MSE, LOO-NLFD and LOO-CBPS respectively.

3 Generalized cross-validation [1]
- A version of LOO-MSE that takes the heterogeneity of the design into account

4 Kernel alignment [2]
- Aligns the eigenvector related to the highest eigenvalue of $K_0$ with the data
- Can also be seen as a similarity between $K_0$ and the covariance matrix obtained from the kernel $K(x, y) = f(x) f(y)$

5 Numerical study
- We use a set of 36 problems
  - Goldstein-Price ($d \in \{1, 2\}$)
  - Mystery ($d = 2$)

Influence of the selection criteria
- We focus on two subsets of problems with different smoothness.
- Fig. 1: 5-dimensional Toms 829 problems.
- Fig. 2: 5-dimensional Rosenbrock and Borehole.
- We compare log $\mathcal{S}$ normalized by "Best" values both with automatically selected and fixed $p$ (Fig. 4: interval score) [4] defined by

$$ \frac{1}{N} \sum_{i=1}^{N} \mathcal{S}(P(z_i), P(\mathbf{x}_i)) = (z_i - \mathbb{E}(z_i))^2 + \frac{1}{N} \sum_{i=1}^{N} \mathbb{H}(z_i) - \mathbb{H}(Y(z_i)). $$

6 Conclusions
- The regularity parameter has a strong impact on the goodness of fit.
- We recommend selecting the regularity from data instead of fixing it to a “standard” value.
- The choice of a reasonable selection procedure has second-order impact but ML and LOO CRPS seem to give the best performances.
- All procedures have the same numerical complexity, using appropriate computations of the selection criteria and their gradients.

References