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 Id: hal-03018559
http://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. Petit1,2 & Julien Beck1 & Paul Feliot2 & Emmanuel Vazquez1
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 $f: X \subset \mathbb{R}^d \to \mathbb{R}$ with Gaussian processes.
- Given data $D_n = \{(x_i, f(x_i))\}$, a Gaussian process $\xi$ can be used to make probabilistic predictions of $f$.

$$\xi(x_i) | D_n \sim N(\xi(x_i); \sigma^2(x_i))$$

- $\xi$ 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.
- Very often, the Matérn covariance function 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 $-\frac{1}{2} \sum l_k^2 + \log(\det(\phi))$

where $\phi_k$ is the covariance matrix of $\xi$ at points $x_k = (x_k, \ldots, x_k)$ for parameters $\theta$ and $l = (x_1, \ldots, x_n)^T$ denotes the values of $f$ at $x_k$.

2 Cross-validation

Leave-one-out (LOO) [3] is a 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 $S: (\mathbb{P}, \mathbb{R}) \to \mathbb{R}$ where $\mathbb{P}$ is a class of probability distributions, with $S(\mathbb{P}, x)$ representing a loss for observing $x$ while predicting $P$.
- Given a scoring rule $S$ the corresponding LOO criterion is

$$L^S(\hat{\theta}) = \sum_{i=1}^n S(N(\hat{\theta}_i, \sigma^2_{\hat{\theta}_i}, x_i), x_i)$$

In this work we consider the following scoring rules [4]

- $S_{\text{gauss}}(P, x) = (\mathbb{E}_{\xi}(x) - x)^2$.
- $S_{\text{kull}}(P, x) = -\log(p(x))$, with $p$ the pdf of $P$.
- $S_{\text{mspe}}(P, x) = \|F - f\|_{L_2^K(R^K)}^2$, with $F$ the cdf of $P$.

We shall denote the resulting selection procedures by LOO-MSE, LOO-NLFD and LOO-CRPS 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 eigenspace related to the highest eigenvalue of $\phi$ with the data.

Can also be seen as a similarity between $\phi$ and the covariance matrix obtained from the kernel $\kappa(x, y) = f(x) f(y)$.

5 Numerical study

We use a set of 36 problems.
- Goldstein-Price of $\xi \in \{1, 2\}$.
- Mystery ( $d = 2$).
- Time $\text{S}^{2,9} \chi^2 \left( k \in \{0, 1, 2\}, d \in \{2, 5\} \right)$.
- Rotated Rosenbrock $\left( d \in \{2, 5\} \right)$.
- Boraborde ($d = 8$) with space-filling designs $X_o \in \{10d, 20d, 50d\}$. For each case:
  - We compare model selection procedures using predictions evaluated on a dense test grid.
  - In particular, we study the influence of the regularity parameter $\rho$ of a Matérn covariance function, by setting $\rho \in \{0, 1, 2, 3, 4, 6, 2d, 2d + \infty\}$ or automatically selecting its value using the selection criteria.
  - We present averaged results through repetitions (using random $X_o$ for instance).

An example

Observe that
- Procedure ML, LOO-MSE, LOO-NLFD and LOO-CRPS give similar accuracies.
- The influence of $\rho$ on the accuracy is strong.

Influence of the selection criteria

We compare the selection procedures with automatically selected $\rho$. Fig. 1: $\text{gauss}^S$ normalized by "Best" values. Fig. 4: interval score [4] defined by

$$\text{SS}_N^{2, d}(u, \alpha, \beta) = \left( u - \beta \right) + \frac{\beta}{\alpha} - 2 \left( \beta - u \right)$$

$\text{with}s^p (x) = \frac{1}{\sqrt{N}} \sum_{i=1}^N s_i^p(x)$

<table>
<thead>
<tr>
<th>$\text{ML}$</th>
<th>$\text{LOO-MSE}$</th>
<th>$\text{LOO-NLFD}$</th>
<th>$\text{LOO-MSPE}$</th>
<th>$\text{LOO-CRPS}$</th>
<th>$\text{KA}$</th>
<th>$\text{OHN}$</th>
<th>$\text{Best}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4846</td>
<td>1.26</td>
<td>0.24</td>
<td>0.14</td>
<td>0.0558</td>
<td>0.071</td>
<td>0.033</td>
<td>0.0558</td>
</tr>
<tr>
<td>0.4846</td>
<td>1.26</td>
<td>0.24</td>
<td>0.14</td>
<td>0.0558</td>
<td>0.071</td>
<td>0.033</td>
<td>0.0558</td>
</tr>
<tr>
<td>0.4846</td>
<td>1.26</td>
<td>0.24</td>
<td>0.14</td>
<td>0.0558</td>
<td>0.071</td>
<td>0.033</td>
<td>0.0558</td>
</tr>
<tr>
<td>0.4846</td>
<td>1.26</td>
<td>0.24</td>
<td>0.14</td>
<td>0.0558</td>
<td>0.071</td>
<td>0.033</td>
<td>0.0558</td>
</tr>
<tr>
<td>0.4846</td>
<td>1.26</td>
<td>0.24</td>
<td>0.14</td>
<td>0.0558</td>
<td>0.071</td>
<td>0.033</td>
<td>0.0558</td>
</tr>
<tr>
<td>0.4846</td>
<td>1.26</td>
<td>0.24</td>
<td>0.14</td>
<td>0.0558</td>
<td>0.071</td>
<td>0.033</td>
<td>0.0558</td>
</tr>
</tbody>
</table>

Table: Average MSE on the validation sets for the different selection procedures and regularity choices.

References

Influence of the selection criteria on the MSE.

Influence of the selection criteria on the interval score.

Fig. 1: Influence of the selection criteria on the MSE.

Fig. 4: Influence of the selection criteria on the interval score.

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