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://hal-centralesupelec.archives-ouvertes.fr/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 Becq1 & Paul Feliot2 & Emmanuel Vazquez1

1 Université Paris-Saclay, CentraleSupélec, Laboratoire des Signaux et Systèmes, Gif-sur-Yvette, France.
2 Safran 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 $D_n = (X_n, f(X_n))$, a Gaussian process $\xi$ can be used to make probabilistic predictions of $f$

$$E[\xi(x)] = \mathbb{E}[f(x)]$$

1 Numerical study

1.1 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

$$\frac{1}{2} \log K_\theta + \text{log det}(K_\theta),$$

where $K_\theta$ is the covariance matrix of $\xi$ at points $X_n = (x_1, \ldots, x_n)$ for parameters $\theta$ and $z = (z_1, \ldots, z_n)^T$ denotes the values of $f$ at $X_n$

1.2 Cross-validation

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

$$S(P, z) = \sum_{i=1}^{n} S(N(\xi(x_i|\theta), \sigma), z_i),$$

where $N(\xi(x_i|\theta), \sigma)$ denotes LOO predictive distributions

In this work we consider the following scoring rules [4]:
- $S_{\text{KL}}(P, z) = \text{KL}(N(z|\xi(x), \sigma) || P)$
- $S_{\text{calc}}(P, z) = -\text{calc}(z, P)$, with $\text{calc}(z, P)$ the pth of $P$
- $S_{\text{calc}}(P, z) = ||F - \mathbb{E}_z[N(\xi(x_i|\theta), \sigma)]||_F$ with $F$ the cdf of $P$

We shall denote the resulting selection procedures by LOO-MSPE, LOO-NLPD and LOO-CRPS respectively.

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

4 Kernel alignment [2]
- Aligns the eigenspace related to the highest eigenvalue of $K_\theta$ with the data
- Can also be seen as a similarity between $K_\theta$ and the covariance matrix obtained from the kernel $\xi(x) \Rightarrow f(x)$

5 Numerical study

We use a set of 36 problems
- Goldstein-Price (of $[1, 2]$)
- Mystery (of $[1, 2]$)
- Time 829 $C^2$ ($k \in [0, 1, 2], d \in [0, 2]$)
- Rotate Rosenbrock ($d \in [0, 2]$)
- Borehole ($d = 8$)

Influence of the selection criteria

We compare the selection procedures with automatically selected $\theta$. Fig. 1 and 2: average LOO-CRPS and ROO-CRPS in each regularity class.

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

<table>
<thead>
<tr>
<th>Metric</th>
<th>ML</th>
<th>LLO-MSPE</th>
<th>LOO-NLPD</th>
<th>LOO-CRPS</th>
<th>GCV</th>
<th>CRPS</th>
<th>KA</th>
<th>OVN</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>1.00</td>
<td>1.12</td>
<td>1.07</td>
<td>1.06</td>
<td>2.16</td>
<td>1.04</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>Parameter</td>
<td>2.00</td>
<td>1.17</td>
<td>1.21</td>
<td>1.29</td>
<td>1.14</td>
<td>1.59</td>
<td>1.66</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Parameter</td>
<td>3.00</td>
<td>1.54</td>
<td>1.80</td>
<td>1.81</td>
<td>1.62</td>
<td>1.36</td>
<td>2.12</td>
<td>1.36</td>
<td>1.36</td>
</tr>
<tr>
<td>Parameter</td>
<td>4.00</td>
<td>2.30</td>
<td>2.32</td>
<td>2.30</td>
<td>2.11</td>
<td>1.21</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>Parameter</td>
<td>5.00</td>
<td>2.36</td>
<td>3.05</td>
<td>2.93</td>
<td>2.66</td>
<td>1.12</td>
<td>4.00</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>Parameter</td>
<td>6.00</td>
<td>2.40</td>
<td>3.30</td>
<td>3.04</td>
<td>2.97</td>
<td>1.01</td>
<td>4.11</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Parameter</td>
<td>7.00</td>
<td>2.66</td>
<td>3.42</td>
<td>3.35</td>
<td>3.16</td>
<td>2.43</td>
<td>4.31</td>
<td>2.43</td>
<td>2.43</td>
</tr>
<tr>
<td>Parameter</td>
<td>8.00</td>
<td>3.22</td>
<td>3.62</td>
<td>3.55</td>
<td>3.16</td>
<td>3.16</td>
<td>4.31</td>
<td>3.16</td>
<td>3.16</td>
</tr>
</tbody>
</table>

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

Influence of the regularity

We focus on two subsets of problems with different smoothness.

Fig. 1: 5-dimensional Time 829 problems; Fig. 2: 5-dimensional Rosenbrock and Borehole

We compare log CRPS normalized by ‘Best’ values with both automatically selected or fixed $\theta \in [0, 1, 2, 3, 4, d, 2d, 4d]$. 

Fig. 3: Influence of the selection criteria on the MSPE

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

References