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
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

\[ \sum_{i=1}^{N} |\mathbf{K}_{\mathbf{X}_i} \mathbf{X}_i + \text{det}(\mathbf{K})| \]

where \( \mathbf{K} \) is the covariance matrix of \( \xi \) at points \( \mathbf{X}_i = (x_1, \ldots, x_d) \) for parameters \( \theta \) and \( z = (z_1, \ldots, z_d)^T \) denotes the values of \( f \) at \( \mathbf{X}_i \).

2 Cross-validation

Leave-one-out (LOO) [5] 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 \( S: \mathcal{P} \times \mathbb{R} \rightarrow \mathbb{R} \) where \( \mathcal{P} \) is a class of probability distributions, with \( S(p, x) \) representing a loss for observing \( x \) while predicting \( p \).

Given a scoring rule \( S \) the corresponding LOO criterion is

\[ L^S_{\text{LOO}}(\hat{\theta}) = \sum_{i=1}^{N} S(N(X_{-i}, \hat{\theta}_{-i}), z_i), \]

where \( N(X_{-i}, \hat{\theta}_{-i}) \) denotes LOO predictive distributions.

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

- \( S_{\text{syn}}(p, z) = \mathbb{E}(\mathbf{K}_{\mathbf{X}_i} \mathbf{X}_i + \text{det}(\mathbf{K})) \)
- \( S_{\text{syn}}(p, z) = \mathbb{E}(\mathbf{K}_{\mathbf{X}_i} \mathbf{X}_i + \text{det}(\mathbf{K})) \), with \( S \) the pdf of \( P \)
- \( S_{\text{syn}}(p, z) = \mathbb{E}(\mathbf{K}_{\mathbf{X}_i} \mathbf{X}_i + \text{det}(\mathbf{K})) \), with \( F \) the cdf of \( P \)

We shall denote the resulting selection procedure by LOO-MSE, LOO-NLPD 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 eigenvector related to the highest eigenvalue of \( \mathbf{K} \) with the data

Can also be seen as a similarity between \( \mathbf{K} \) and the covariance matrix obtained from the kernel (\( s \), \( t \)) \( \rightarrow f(s,t) \).

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 \( \hat{R}_{\text{AIC}} \) normalized by "Best" values; Fig. 4: interval score (4) defined by

\[ s_{\text{LOO}}(p, z) = (s - l) + \frac{2}{n} \left( 1 - \frac{1}{n} \right) \]

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 [6]

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

<table>
<thead>
<tr>
<th>MSE</th>
<th>ML</th>
<th>LOO-MSE</th>
<th>LOO-NLPD</th>
<th>LOO-CRPS</th>
<th>KA</th>
<th>GCV</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.08</td>
<td>1.12</td>
<td>1.07</td>
<td>1.06</td>
<td>2.16</td>
<td>1.04</td>
<td>1.07</td>
</tr>
<tr>
<td>25</td>
<td>1.15</td>
<td>1.11</td>
<td>1.06</td>
<td>1.04</td>
<td>2.14</td>
<td>1.04</td>
<td>1.01</td>
</tr>
<tr>
<td>50</td>
<td>1.17</td>
<td>1.21</td>
<td>1.09</td>
<td>1.07</td>
<td>2.16</td>
<td>1.04</td>
<td>1.01</td>
</tr>
<tr>
<td>100</td>
<td>1.24</td>
<td>1.80</td>
<td>1.07</td>
<td>1.05</td>
<td>2.16</td>
<td>1.04</td>
<td>1.01</td>
</tr>
<tr>
<td>200</td>
<td>1.40</td>
<td>2.32</td>
<td>1.05</td>
<td>1.07</td>
<td>2.16</td>
<td>1.04</td>
<td>1.01</td>
</tr>
<tr>
<td>400</td>
<td>2.36</td>
<td>1.05</td>
<td>1.07</td>
<td>2.16</td>
<td>1.04</td>
<td>1.01</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Influence of the regularity on the loss for smooth problems

Fig. 2. Influence of the regularity on the loss for non-smooth problems

Fig. 3. Influence of the selection criteria on the MSE

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

References


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