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
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

1 Maximum-likelihood [5]

A very popular technique

Choose the parameters that yield the highest value of the probability density function (pdf) of the observations, or equivalently, minimize

$$L(\theta) = -\frac{1}{2} \sum_{i=1}^{N} (y_i - f(x_i))^2$$

where $y_i$ are the observations, $f(x_i)$ the pdf of the Gaussian process at points $x_i$.

2 Cross-validation

Leave-one-out (LOO) [6] 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 $(P, z) \rightarrow R \rightarrow R$ 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:

$$SLOO(x_i) = \frac{1}{N} \sum_{j \neq i} S(N(x_i, y_j, \hat{f}(x_i), z_j), z_i)$$

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 ($x, y \rightarrow f(x,y)$).

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 compare the selection procedures with automatically selected $p$, $\text{log}(\text{Sagitt})$ normalized by ‘Best’ values.

$$SLOO(x_i) = \frac{1}{N} \sum_{j \neq i} S(N(x_i, y_j, \hat{f}(x_i), z_j), z_i)$$

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

- Tong 829 [4] (d ∈ {0, 1, 2, 3, 4, 6, 2d, 2d + 4}) or automatically selecting its value using the selection criteria.

4.2 Influence of the selection criteria on the MSE.

- Gaussian processes
- Generalized cross-validation
- Cross-validation
- Bootstrap

with space-filling designs $X_n$ ∈ {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 $p$ of a Gaussian process covariance function, by setting $p \in \{0, 1, 2, 3, 4, 6, 2d, 2d + 4\}$ or automatically selecting its value using the selection criteria.

We present averaged results through repetitions (using random $X_n$ for instance).

An example

Observe that:

- Procedure ML, LOO-MSE, LOO-NLPD and LOO-CRPS give similar accuracies.
- The influence of $p$ on the accuracy is strong.

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>CRPS</th>
<th>LOO-CRPS</th>
<th>KA</th>
<th>OVN</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.88 10 -2</td>
<td>1.12 10 -2</td>
<td>1.07 10 -2</td>
<td>1.00 10 -2</td>
<td>0.26 10 -2</td>
<td>0.14 10 -2</td>
<td>0.04 10 -2</td>
<td>0.74 10 -2</td>
</tr>
<tr>
<td>0</td>
<td>1.17 10 -2</td>
<td>1.44 10 -2</td>
<td>1.28 10 -2</td>
<td>0.85 10 -2</td>
<td>0.71 10 -2</td>
<td>0.27 10 -2</td>
<td>0.18 10 -2</td>
<td>0.74 10 -2</td>
</tr>
<tr>
<td>0</td>
<td>1.17 10 -2</td>
<td>1.44 10 -2</td>
<td>1.28 10 -2</td>
<td>0.85 10 -2</td>
<td>0.71 10 -2</td>
<td>0.27 10 -2</td>
<td>0.18 10 -2</td>
<td>0.74 10 -2</td>
</tr>
<tr>
<td>0</td>
<td>1.17 10 -2</td>
<td>1.44 10 -2</td>
<td>1.28 10 -2</td>
<td>0.85 10 -2</td>
<td>0.71 10 -2</td>
<td>0.27 10 -2</td>
<td>0.18 10 -2</td>
<td>0.74 10 -2</td>
</tr>
<tr>
<td>0</td>
<td>1.17 10 -2</td>
<td>1.44 10 -2</td>
<td>1.28 10 -2</td>
<td>0.85 10 -2</td>
<td>0.71 10 -2</td>
<td>0.27 10 -2</td>
<td>0.18 10 -2</td>
<td>0.74 10 -2</td>
</tr>
<tr>
<td>0</td>
<td>1.17 10 -2</td>
<td>1.44 10 -2</td>
<td>1.28 10 -2</td>
<td>0.85 10 -2</td>
<td>0.71 10 -2</td>
<td>0.27 10 -2</td>
<td>0.18 10 -2</td>
<td>0.74 10 -2</td>
</tr>
<tr>
<td>0</td>
<td>1.17 10 -2</td>
<td>1.44 10 -2</td>
<td>1.28 10 -2</td>
<td>0.85 10 -2</td>
<td>0.71 10 -2</td>
<td>0.27 10 -2</td>
<td>0.18 10 -2</td>
<td>0.74 10 -2</td>
</tr>
</tbody>
</table>

References


[8] Currin, C. Interpolation of Spatial Data: Some Theory for Kriging