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 Beck1 & Paul Feliot2 & Emmanuel Vazquez1
1Université 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 $f: \mathbb{R} \rightarrow \mathbb{R}$ with Gaussian processes
- Given data $D_n = \{X_n, f(X_n)\}$, a Gaussian process $\xi$ can be used to make probabilistic predictions of $f$

\[
\xi(x) | D_n \sim \mathcal{N}(\xi(x), \sigma^2(x))
\]

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

\[
L(M) = \log p(D_n | M)
\]

where $M$ is the covariance matrix of $\xi$ at points $X_n = \{x_1, \ldots, x_n\}$ for parameters $\theta$ and $x = (x_1, \ldots, x_n)^T$ denotes the values of $f$ at $X_n$

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 $S: (\mathbb{R}, \mathbb{R}) \rightarrow \mathbb{R}$ where $P$ 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 LOO criterion is

\[
\frac{1}{n} \sum_{i=1}^{n} S(\hat{N}(x_i | \theta_{n-1}, \sigma^2_{n-1}), z_i)
\]

where $N(\cdot | \cdot, \sigma^2)$ denotes LOO predictive distributions.
In this work we consider the following scoring rules [4]:
\[
S_{\text{Brier}}(P, x) = (\hat{y}_x - x)^2
\]
\[
S_{\text{SIP}}(P, x) = -\log(p(x|P))
\]
\[
S_{\text{Shapiro}}(P, x) = \| F - \log p(F|\xi) \|^2
\]
We shall denote the resulting selection procedures by LOO-MSPE, LOO-NLFD and LOO-CRPS respectively.

3 Generalized cross-validation [1]
- A version of LOO-MSPE 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 $k(x, y) \rightarrow (x)(y)$

5 Numerical study
We use a set of 16 problems:
- Goldstein-Price ($d \in \{1, 2\}$)
- Mystery ($d = 2$)

Influence of the selection criteria
We compare the selection procedures with automatically selected $\theta$, Fig. 1: log $\hat{S}_{\text{Brier}}$ normalized by ‘Best’ values, Fig. 4: interval score [4] defined by

\[
\frac{\hat{S}_{\text{Brier}}(x, u, z)}{\text{Best}} = \left( \hat{y}_x - f(x) \right) \frac{\hat{y}_x - f(x)}{\hat{y}_u - \hat{y}_z}
\]

6 Conclusions
- The estimator parameter $\theta$ has a strong impact on the goodness of fit
- We recommend selecting the regularization from data instead of fixing it to a ‘standard’ value
- The choice of a reasonable selection procedure has second-order impact but 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 MSPE on the validation sets for different selection procedures and regularity choices.

Influence of the regularity
We focus on two subsets of problems with different smoothness: Fig. 1: 5-dimensional Tors 829 problems, Fig. 2: 5-dimensional Rosenbrock and Borehole.
We compare log $\hat{S}_{\text{Brier}}$ normalized by ‘Best’ values both with automatically selected or fixed $\theta$ in $[0, 1, 2, 3, 4, 6, 2d, 2d \cdot 4\alpha]$.

References

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

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

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

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