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\textsuperscript{1,2} & Julien Bec\textsuperscript{1} & Paul Feliot\textsuperscript{2} & Emmanuel Vazquez\textsuperscript{1}
\textsuperscript{1}Université Paris-Saclay, CentraleSupélec, Laboratoire des Signaux et Systèmes, Gif-sur-Yvette, France.
\textsuperscript{2} Safran Aircraft Engines, Moissy-Cramayel, France.

email: sebastien.petit@centralesupelec.fr

Context

\begin{itemize}
  \item Exploration of black-box numerical simulations $f: \mathbb{R}^d \to \mathbb{R}$ with Gaussian processes.
  \item Given data $D_0 = (X_0, f|X_0)$, a Gaussian process $\xi$ can be used to make probabilistic predictions of $f$.
\end{itemize}

\begin{equation}
\mathbb{E}(x|D_0) = \mathbb{N}(\hat{\mu}(x), \hat{\sigma}(x))
\end{equation}

\begin{itemize}
  \item $\hat{\mu}$ is a prior over functions.
  \item The choice of $\xi$ is critical for good predictions and design-of-experiments techniques.
\end{itemize}

The prior $\xi$ is often chosen within a parametric family.

- Very often, the Matérn covariance function family 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]

\begin{itemize}
  \item A very popular technique.
  \item Choose the parameters that yield the highest value of the likelihood density for the observations, or equivalently, minimize

$$L^2(K_0) = \sum_{i=0}^{n} \mathbb{N}(x_i|K_0),$$

where $K_0$ is the covariance matrix of $\xi$ at points $x = (x_1, \ldots, x_n)$ for parameters $\theta$ and $z = (z_1, \ldots, z_n)$ denotes the values of $f$ at $X_0$.

2 Cross-validation

Leave-one-out (LOO) [5] is a second very popular technique.

\begin{itemize}
  \item Consists in averaging losses for predicting one observation using the others.
  \item We suggest using negatively-oriented scoring rules [4] for the loss functions.
  \item A (negatively-oriented) scoring rule is a mapping $S: (P, R) \to R$ where $P$ is a class of probability distributions, with $S(P, z)$ representing a loss for observing $z$ while predicting $P$.
  \item Given a scoring rule $S$ the corresponding LOO criterion is

$$L^2(S, \hat{\theta}) = \sum_{i=0}^{n} S(N(\hat{\mu}(x_i|K_{\hat{\theta}}), \hat{\sigma}(x_i)), z_i).$$

\end{itemize}

\begin{equation}
S(N(\hat{\mu}(x_i|K_{\hat{\theta}}), \hat{\sigma}(x_i)), z_i)
\end{equation}

3 Generalized cross-validation [1]

\begin{itemize}
  \item A version of LOO-MSE that takes the heterogeneity of the design into account.
\end{itemize}

4 Kernel alignment [2]

\begin{itemize}
  \item Aligns the eigenspace related to the highest eigenvalue of $K_0$ with the data.
  \item Can also be seen as a similarity between $K_0$ and the covariance matrix obtained from the kernel $\langle x, y \rangle \to f(x)g(y)$.
\end{itemize}

5 Numerical study

\begin{itemize}
  \item We use a set of 36 problems:
    \begin{itemize}
      \item Goldstein-Price ($d \in \{1, 2\}$).
      \item Mystery ($d = 2$).
    \end{itemize}
\end{itemize}

Influence of the selection criteria

We compare the selection procedures with automatically selected $p$.

- For $\hat{\theta} = \text{Best}$ values, Fig. 4: interval score normalized by "Best" values.

\begin{equation}
\frac{\text{CRPS}(\hat{\theta})}{\text{CRPS}(\text{Best})}
\end{equation}

6 Conclusions

\begin{itemize}
  \item The regularity parameter has a strong impact on the goodness of fit.
  \item We recommend selecting the regularity from data instead of fixing it to a "Standard" value.
  \item The choice of a reasonable selection procedure has second-order impact but ML and LOO CRPS seem to give the best performances.
  \item All procedures have the same numerical complexity, using appropriate computations of the selection criteria and their gradients [6].
\end{itemize}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
Procedure & ML & LOO-MSE & LOO-NLDF & CGT-CRPS & KA & CRPS \hline
\hline
Random & 0.04 & 3.72 & 1.22 & 0.29 & 0.15 & 0.03 \hline
CRPS & 0.04 & 3.72 & 1.22 & 0.29 & 0.15 & 0.03 \hline
\end{tabular}
\caption{Average MSPE on the validation sets for the different selection procedures and regularity choices.}
\end{table}

References