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

\[ \xi(f) \sim N(\xi(f), \sigma_\xi(f)) \]

(1)

\( \xi \) is a prior over functions.

The choice of \( \xi \) is critical for good predictions and design-of-experiments techniques.

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

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]

A very popular technique

Choose the parameters that yield the highest value of the probability density for the observations, or equivalently, minimize

\[ \sum_{i=1}^{n} |y_i - f(x_i)|^p \]

where \( K_\theta \) is the covariance matrix of \( \xi \) at points \( x = (x_1, \ldots, x_n) \) for parameters \( \theta \) and \( z = (z_1, \ldots, z_m) \) denotes the values of \( f \) at \( x_0 \).

2 Cross-validation

Leave-one-out (LOO) [4] 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 \( S : (\mathbb{P}, R) \rightarrow R \) where \( \mathbb{P} \) is a class of probability distributions, with \( S(P) \) representing a loss for observing \( z \) while predicting \( P \).

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

\[ L_{\text{LOO}}(\theta) = \frac{1}{n} \sum_{i=1}^{n} S(\{y_i, \ldots, y_n, z_i \}) \]

where \( N(z_i | \mathcal{K}_\theta, \sigma^2(z_i)) \) denotes LOO predictive distributions.

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

\[ S_{\text{MSE}}(P) = \mathbb{E}[(y_i - f_i)^2] \]

\[ S_{\text{MSPE}}(P) = \mathbb{E}[(y_i - f_i)^2 | f_0] \]

\[ S_{\text{SSC}}(P) = \mathbb{E}[(F - \mathbb{E}[F | x_i])^2] \]

where \( F \) is the forecast and \( x_i \) is the input.

We shall denote the resulting selection procedure by LOO-MSE, LOO-MSPE and LOO-SSC 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 \( K_\theta \) with the data.

Can also be seen as a similarity between \( K_\theta \) and the covariance matrix obtained from the kernel \( (x, y) \rightarrow f(x)g(y) \)

5 Numerical study

We use a set of 36 problems:

- Goldstein-Price \( \xi \in [1, 2] \)
- Mystery \( (d = 2) \)
- 450-dimensional T09 problems; Fig. 2:
- Gaussian process model selection for computer experiments

Sébastien J. Petit, Julien Bect, and Paul Feliot.

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

References


