Cross Validation and Maximum Likelihood estimations of hyperparameters of Gaussian processes with model misspecification

François Bachoc1, Joséfin Garnier1 and Jean-Marc Martinez1

1CEA-Saclay, DENS, DM2S, SFMIE, LIGS, F-91191 Gif-sur-Yvette, France.
2Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VII

Introduction

- Main Context: Estimation of the variance hyperparameter in Kriging in case of Model misspecification.
- Goals:
  - Comparison of Maximum Likelihood (ML) and Cross Validation (CV) in case of model misspecification.
  - Comparison based on Predictive Variance reliability

Framework

- Observation of a centered, unit variance stationary Gaussian process \( Y \) on \( X \) with covariance function \( C_{1} \).
- Vector \( y \) of observations on \( x_{1}, \ldots, x_{n} \in X \).
- Kriging metamodel \( y_{k} = (y_{0}, \hat{\sigma}^{2}(\sigma_{k})^{2}) \) given by the set \( C \) of covariance functions:
  \[ C = \{ \tilde{\sigma}^{2}C_{2}, \sigma \in \mathbb{R}^{+} \} \]
  with \( C_{2} \) stationary correlation function, \( C_{2} \neq C_{1} \) model misspecification
- Quantity of interest for \( \sigma \): the Risk at \( x_{0} \)
  \[ R_{x_{0}} = \mathbb{E} \left[ (y_{0} - y_{\hat{\sigma}}^{2} - \hat{\sigma}^{2}(\sigma_{k})^{2}) \right] \]

Analytical expression of the risk for an estimator \( \hat{\sigma}^{2} \) of the form \( \hat{\sigma}^{2}M_{y} \)
\[ R_{x_{0}} = f(M_{y}M_{0}) + 2r_{1}(M_{y}) - 2r_{2}(M_{y}, M_{0}) \]
\[ \tilde{\sigma}^{2} = 2r_{1}(\hat{\sigma}(M_{y}) + r_{2}(M_{y}, M_{0})), \]

With:
\[ f(A, B) = tr(A)(tr(B) + 2tr(AB)) \text{ for } A, B \in \mathbb{R}^{n \times n} \text{ real symmetric matrices}, \]
\[ M_{y} = \mathbf{R}_{L}^{2}(||r_{i}||_{2})^{-1}, \quad R_{L}^{2}(||r_{i}||_{2})^{-1}, \quad r_{L}^{2}(||r_{i}||_{2})^{-1} \]
\[ c_{1} = 1 - r_{L}^{2}(||r_{i}||_{2}), \quad c_{2} = 1 - r_{L}^{2}(||r_{i}||_{2}) \]

CV and ML estimation

- Maximum Likelihood (ML) estimator:
  \[ \hat{\sigma}^{2}_{ML} = \frac{1}{n} \sum_{i=1}^{n} \left(y_{i} - \hat{y}_{i}^{2}ight)^{2} \]
- Cross Validation (CV) estimator:
  \[ \hat{\sigma}^{2}_{CV} = \frac{1}{n} \sum_{i=1}^{n} \left(y_{i} - \hat{y}_{i}^{2}ight)^{2} \]

\[ \hat{y}_{i} = \mathbb{E} \left[ y_{i} \right], \quad \hat{y}_{i} = \mathbb{E} \left[ y_{i} \right], \quad \text{with } \mathbb{E} \left[ y_{i} \right] \text{ based on } (y_{i}, y_{i-1}, y_{i+1}) \text{ with covariance function } C_{0} \]

- Thanks to the virtual Leave One Out form [Dub98] we have:
  \[ \hat{\sigma}^{2}_{CV} = \frac{1}{n} \left(1 - \frac{1}{n} \right) \sum_{i=1}^{n} \left(y_{i} - \hat{y}_{i}^{2}ight)^{2} \]
- Case \( C_{1} = C_{2} \): ML reaches the Charnes-Rau bound [2]
- Case \( C_{1} \neq C_{2} \): numerical evaluation of the risk formulas

Numerical results

- Quantities of interest for an estimator \( \hat{\sigma}^{2} \):
  
<table>
<thead>
<tr>
<th>Quantity of interest</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk on Target Ratio (BTR)</td>
<td>( \text{BTR} \sigma(x) = \frac{\sqrt{\hat{\sigma}^{2}(\sigma_{k})^{2}}}{\hat{\sigma}} )</td>
</tr>
<tr>
<td>Integrated Risk on Target Ratio (IHTR)</td>
<td>( \text{IHTR} = \frac{1}{n} \text{BTR}<em>{\sigma}(\hat{y}</em>{i}) )</td>
</tr>
<tr>
<td>Bias on Target Ratio (BTR)</td>
<td>( \text{BTR}<em>{\sigma} = \frac{\sqrt{\hat{\sigma}^{2}(\sigma</em>{k})^{2}}}{\hat{\sigma}} )</td>
</tr>
<tr>
<td>Integrated Bias on Target Ratio (IBT)</td>
<td>( \text{IBT}<em>{\sigma} = \frac{1}{n} \text{BTR}</em>{\sigma}(\hat{y}_{i}) )</td>
</tr>
</tbody>
</table>

- Procedure: We take \( X = [0, 1]^{d} \) with uniform measure. We generate \( n_{y} \) designs \( (x_{1}, \ldots, x_{n}) \) using the LHS-Maximin technique; compute each time the four criteria above (analytical formulation and Monte Carlo for integration) and plot the average.
- Setting for the figures:
  
<table>
<thead>
<tr>
<th>Function ( n_{x}, n_{y} )</th>
<th>Case for correlation model</th>
<th>MSE</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ishigami</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Morris</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

- When model misspecification (high MSE) CV performs better than ML. Model misspecification can come from a too large model (over-parametrization) for Ishigami or a too small model (under-parametrization) for Morris.

Conclusion

- In our studies: when the model misspecification becomes important, the CV performs better than ML.
- Other Cross Validation criteria?

References