Lab 2

Gaussian Processes.

David Padilla Orenga
Ignacio Pastore Benaim

October 27, 2024

1 Disclaimer

Throughout this lab we have used ChatGPT for the following general tasks:

- Understanding thoroughly the theory behind each task.
- Reviewing and correcting the code.
- Checking misspellings and rewriting sections of this report.
- Brainstorming ideas to comprehensively complete each task.

2 Part 1

First we tested the code with subsamples = 1, as in the default notebook. After implementing the whole code, we tested the models with subsamples = 20.

2.1 Which kernel provide the best fit?

We tested RBF, Cosine and the sum between RBF and Matern52 kernels (called *k_sum*).

The best fit was provided by the sum of RBF and Matern52. This is because the RBF kernel is good at capturing the long-range correlations, while the Matern52 kernel is good at capturing the short-range correlations.

2.2 Would it be possible to get a better fit than by ML?

We made an hyperparameter tuning for each model using the the method "optimize_restarts(num_restarts = 5)".

2.3 Compare the noiseless and "normal" predictions.

In the notebook we implemented both methods for the best model. The election depends on the application. We would choose the noisy model if it would be applied in a real-world task, where the noise of the data would enrich the whole model. Otherwise, if the model would be applied in a simulation, for instance, the noiseless model would be a better choice to make predictions.

3 Part 2

We tested 4 kernels for each contaminant: NO2, O3, NO, NOx. The kernels were: RBF, Matern32, Matern52 and RationalQuadratic.

3.1 Use different kernels to predict the concentration for each of the contaminants.

Based on empirical tests across all contaminants (NO2, O3, NO, NOx), the Radial Basis Function (RBF) kernel consistently yielded the most reliable predictions.

3.2 hyperparameter learning

To address challenges in hyperparameter learning, particularly ensuring convergence to optimal values, we employed an iterative optimization approach using optimize_restarts = 5. This process was conducted separately for each contaminant, with each RBF kernel initialized using averaged parameter values derived from prior optimization runs.

3.3 If you can invest in a new station. Where would you place it?.

To maximize the informational gain from an additional monitoring station, we recommend situating it in an area with consistently high uncertainty (variance) across contaminants. Visual inspection of the variance plots across contaminants suggests that the Casablanca or Parque del Agua regions exhibit higher predictive uncertainty, indicating these locations could most benefit from additional data.

4 Part 3

4.1 Implement prediction and log-likelihood.

We decided to implement both prediction and log-likelihood methods using first, the standard equations and, afterwards the Cholesky decomposition. With the first method we ensured that the code was working properly. Then, we implemented the Cholesky decomposition for the extra point. Final results are presented with the Cholesky decomposition.

4.2 Kernel Comparison for Gaussian Process Regression

We implemented and compared two kernels, the squared exponential and Matérn 3/2, for Gaussian Process regression on the weighdata set. The GP class was structured to allow easy kernel switching, providing flexibility in model experimentation. The dataset was splitted into training and test sets, with 80% of the data used for training and validating, and 20% for testing.

4.2.1 Cross-Validation and Hyperparameter Tuning

To ensure robust evaluation, we applied 5-fold cross-validation, optimizing hyperparameters separately for each fold. For performance metrics, Mean Squared Error (MSE) was used to gauge predictive accuracy, while Negative Log-Likelihood (NLL) evaluated model uncertainty, capturing how well the model fit the data with calibrated confidence. The cross-validation results, averaged across folds, were as follows:

- **Squared Exponential Kernel**: MSE = 0.534 ± 0.077, NLL = 621.13 ± 10.08
- **Matérn 3/2 Kernel**: MSE = 0.513 ± 0.071, NLL = 623.60 ± 9.68

4.2.2 Final Evaluation

After cross-validation, we retrained the GP models on the entire training set, retuned the hyperparameters and evaluated them on the test set to obtain realistic performance estimates.

Test set results were as follows:

- **Squared Exponential Kernel**: Test MSE = 0.438, Test NLL = 770.26
- **Matérn 3/2 Kernel**: Test MSE = 0.438, Test NLL = 770.26

Results showed comparable performance between kernels, with both achieving low MSE and calibrated uncertainty levels. This suggests that both kernels are well-suited for the regression task. More kernels could be tested following the same methodology to further explore model performance.