Filtering Outliers in Bayesian Optimization

Ruben Martinez-Cantin, Kevin Tee, Michael McCourt, Katharina Eggensperger

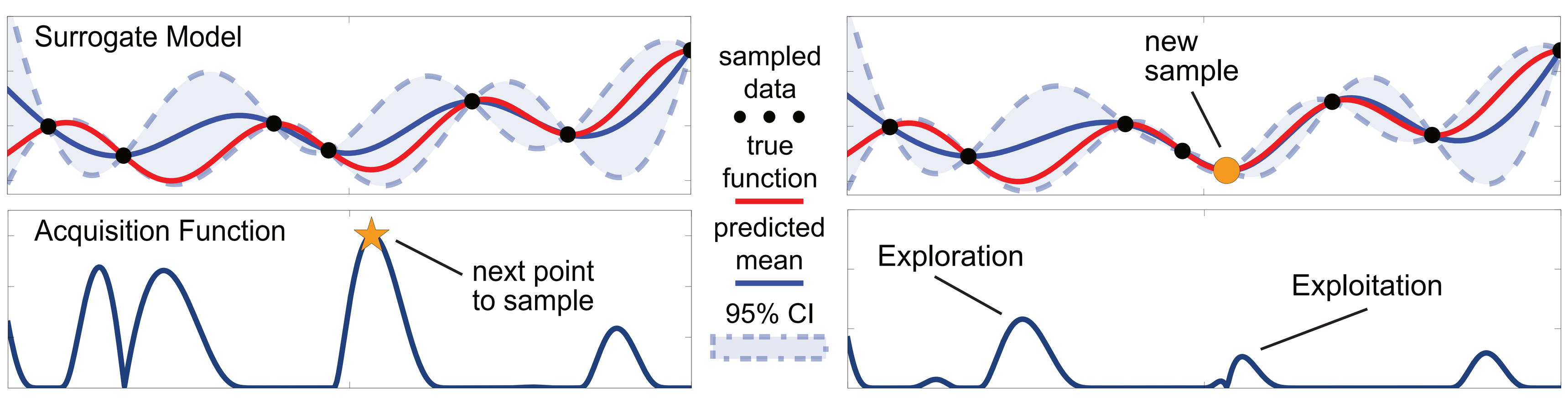
SIGOPT · UNI FREIBURG · Centro Universitario de la Defensa Zaragoza

Bayesian Optimization

Bayesian optimization (BO) is a powerful tool for efficient global optimization of stochastic and nonconvex objectives.

Initial data is sampled from prior beliefs. Then, in subsequent iterations,

- A probabilistic surrogate model is built from the current data, and
- An acquisition function is defined (using the model) and then maximized to identify the most beneficial point to next sample.



Optimization Memory - Part of the power of Bayesian optimization comes from this surrogate model; it incorporates all previously observed data and provides a memory of the optimization progress.

Outliers

The standard assumptions of BO include some observation uncertainty.

$$y_i = f(\mathbf{x}_i) + \epsilon_i, \qquad 1 \leq i \leq N, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_n^2)$$

We define outliers in our context to be a situation where, with probability p, there is an **unexpected**, and **undetected**, deviation from this uncertainty assumption.

Machine learning provides examples of such outliers.

- Scenario - Neural network trained on Boston housing data.
  Outlier - Erroneous early stopping of gradient descent.
- Scenario - Variational autoencoder trained on MNIST data.
  Outlier - I/O errors cause training on a small subset of data.

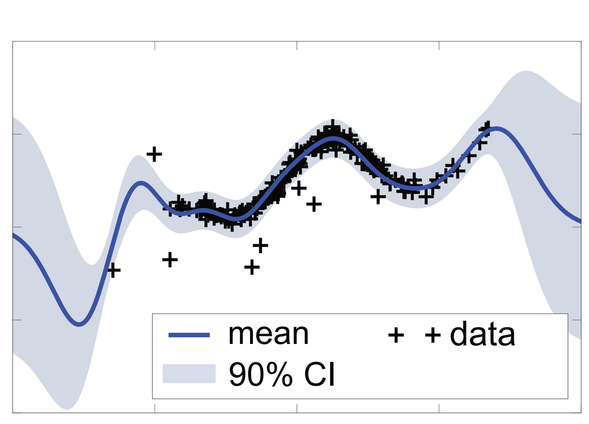
Because these outliers go undetected, the objective values are not recognized immediately as incorrect, leading them to be included in the construction of the surrogate model.

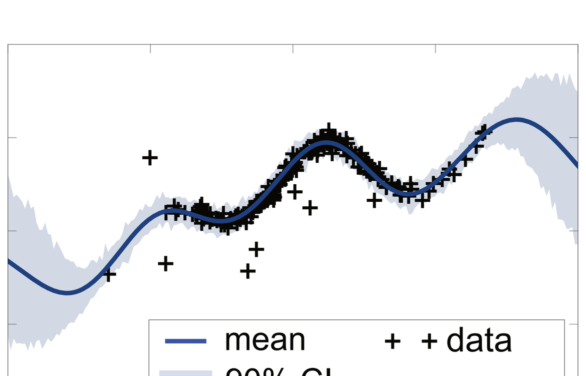
Question - What impact will these outliers have on the surrogate model and subsequent samples?

Robust Regression & Gaussian Processes

Fitting a standard Gaussian process (GP) model to data with outliers can produce:

- unreasonably high variance
- inaccurate predictions



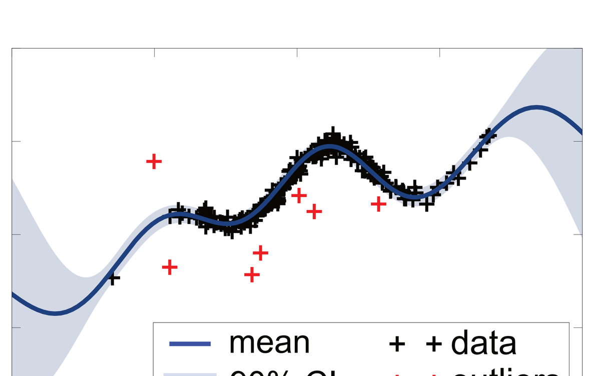


The field of robust regression provides tools for building models containing outliers.

Strategy - Using a Student's-t model for the noise produces better behaved predictions, albeit at greater cost and with less stability.

This more robust model provides a method for filtering outliers.

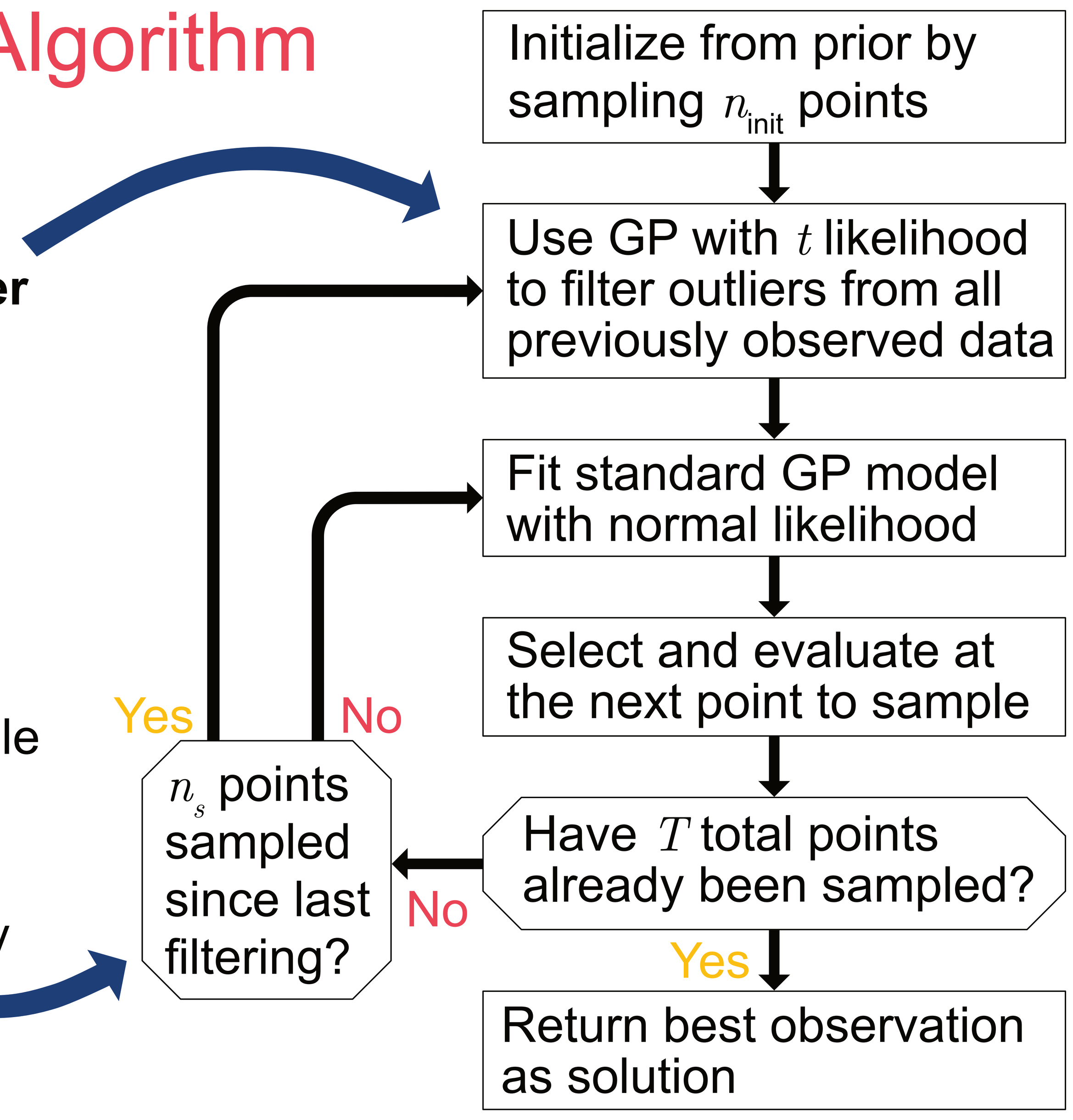
Our approach - Filter outliers from the data and use what remains to build a standard GP model for the remaining points.



Our Filtering Algorithm

Goal - Utilize robust regression tools to **filter outliers** from data and apply standard BO to the remaining points.

Goal - Utilize a schedule to minimize expensive robust model fittings; only filter all data every n_s observations.

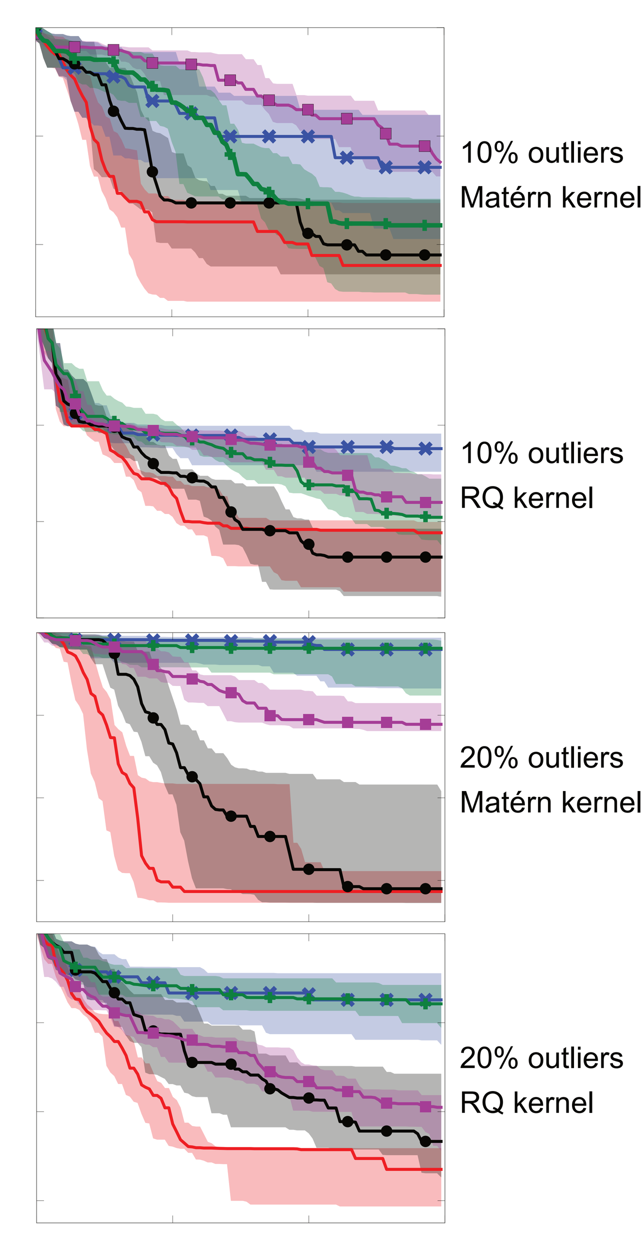


Outcome - Using the t likelihood for filtering (not the next point selection) is more efficient than using a robust model for the BO.

Comparison to Robust Regression

Minimizing synthetic functions drawn from GPs with manufactured outliers allows us to directly compare:

- Our Method
- Baseline (BO with normal likelihood)
- BO with t likelihood
- BO with t process
- No outliers (ideal)

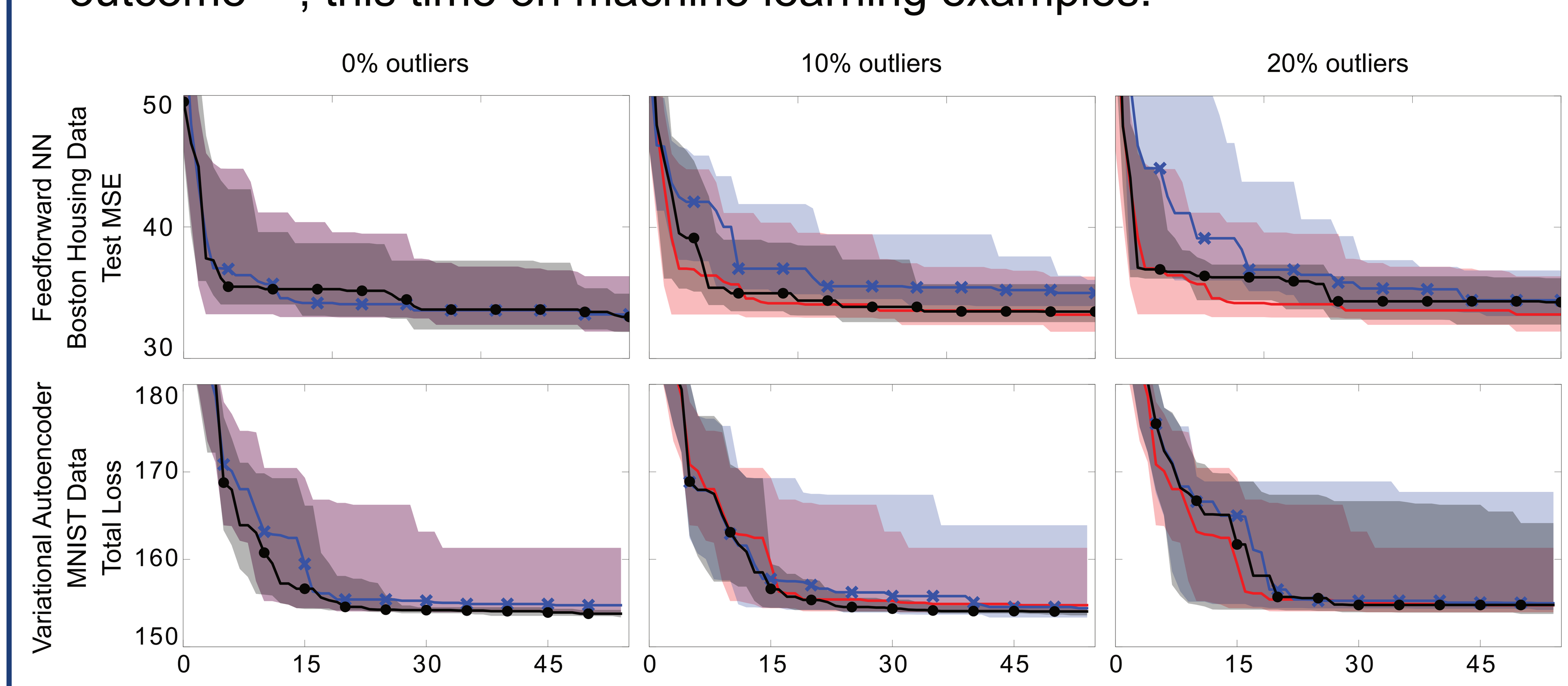


Our takeaways from these results include:

- Ignoring the outliers is infeasible, as the Baseline consistently fails to converge.
- Performing BO with robust regression methods + and ■ produces inconsistent results that are inferior to our method.

Tuning ML Hyperparameters

We again compare our method • against the baseline × and the ideal outcome –, this time on machine learning examples.



Conclusion - Our filtering strategy provides better accuracy and a significant reduction in variance compared to the baseline.

Jarno Vanhatalo, Pasi Jylänki, and Aki Vehtari. Gaussian process regression with Student-t likelihood. In NIPS, pages 1910–1918, 2009.
Amar Shah, Andrew Gordon Wilson, and Zoubin Ghahramani. Student-t processes as alternatives to Gaussian processes. In AISTATS, pages 877–885, 2014.
Anthony O'Hagan. On outlier rejection phenomena in Bayes inference. Journal of the Royal Statistical Society. Series B, pages 358–367, 1979.
Pasi Jylänki, Jarno Vanhatalo, and Aki Vehtari. Robust Gaussian process regression with a Student-t likelihood. JMLR, 12(Nov):3227–3257, 2011.

BayesOpt 2017 Workshop at NIPS, December 9, 2017, Long Beach, CA, USA