Pontificia Universidad Católica de Chile
Instituto de Ingeniería Biológica y Médica
Escuela de Ingeniería

Multi-fidelity classification with Gaussian process: accelerating the prediction of large-scale computational models

Francisco Sahli Costabal[1]*, Paris Perdikaris[2], Ellen Kuhl[3], Daniel Hurtado[1]

[1]Pontificia Universidad Católica de Chile, [2]University of Pennsylvania, [3]Stanford University, *fsahli1@uc.cl

Motivation

Machine learning techniques typically rely on large datasets to create accurate classifiers. However, there are situations when data is scarce and expensive to acquire. This is the case of studies that rely on state-of-the-art computational models which typically take days to run, thus hindering the potential of machine learning tools. Nonetheless, there are usually lower fidelity approximations of the model that are less expensive to compute and can provide valuable information. In recent years, there has been an increased attention in the machine learning community to develop predictive methods that enable the effective fusion of variable fidelity information sources [1]. However, none of these techniques work in the case of binary output. In this work, we present a novel classifier that takes advantage of lower fidelity models and inexpensive approximations to predict the binary output of expensive computer simulations.

Methods

We have data coming from an expensive, high fidelity model (H) and from an inexpensive low fidelity approximation of model (L), organized in input/output label pairs:

$$\mathcal{D} = [\{(\boldsymbol{x}_{L_i}, \hat{y}_{L_i})_{i=1}^{N_L}\}, \{(\boldsymbol{x}_{H_i}, \hat{y}_{H_i})_{i=1}^{N_H}\}] \qquad \hat{y}_i = \{0,1\}$$

We postulate an autoregressive model [1] between the different levels of fidelity with Gaussian process priors:

$$\begin{aligned} \boldsymbol{y}_L &= f_L(\boldsymbol{x}_L) + \epsilon_L \\ \boldsymbol{y}_H &= f_H(\boldsymbol{x}_H) + \epsilon_H \\ f_H(\boldsymbol{x}) &= \rho f_L(\boldsymbol{x}) + \delta(\boldsymbol{x}) \end{aligned} \qquad \begin{aligned} f_L(\boldsymbol{x}) &\sim \mathcal{GP}(0, k_L(\boldsymbol{x}, \boldsymbol{x}'; \theta_L)) \\ \delta(\boldsymbol{x}) &\sim \mathcal{GP}(0, k_H(\boldsymbol{x}, \boldsymbol{x}'; \theta_H)), \end{aligned}$$

We pass the output of the latent functions through a sigmoid function to obtain meaningful probabilities and assign a Bernoulli likelihood to the binary data:

$$\begin{aligned} p(\hat{y}_i = +1) &= \sigma(y_i) \\ \hat{y}_i &\sim \text{Bernoulli}(p) \\ i &= [L, H] \end{aligned}$$

We adopt a fully Bayesian treatment for the hyper-parameters. We use the NO-U-Turn sampler [2], which is part of Hamiltonian Monte Carlo methods. We take advantage of the probabilistic nature of the classifier to implement active learning strategies. We adaptively select the points near the boundary with high variance [3]. We also introduce a sparse approximation [4] to enhance the ability of the classifiers to handle large datasets.

Results

We test these multi-fidelity classifiers against their single-fidelity counterpart with synthetic data (Figures 1 and 2), showing a median computational cost reduction of 23% for a target accuracy of 90%.

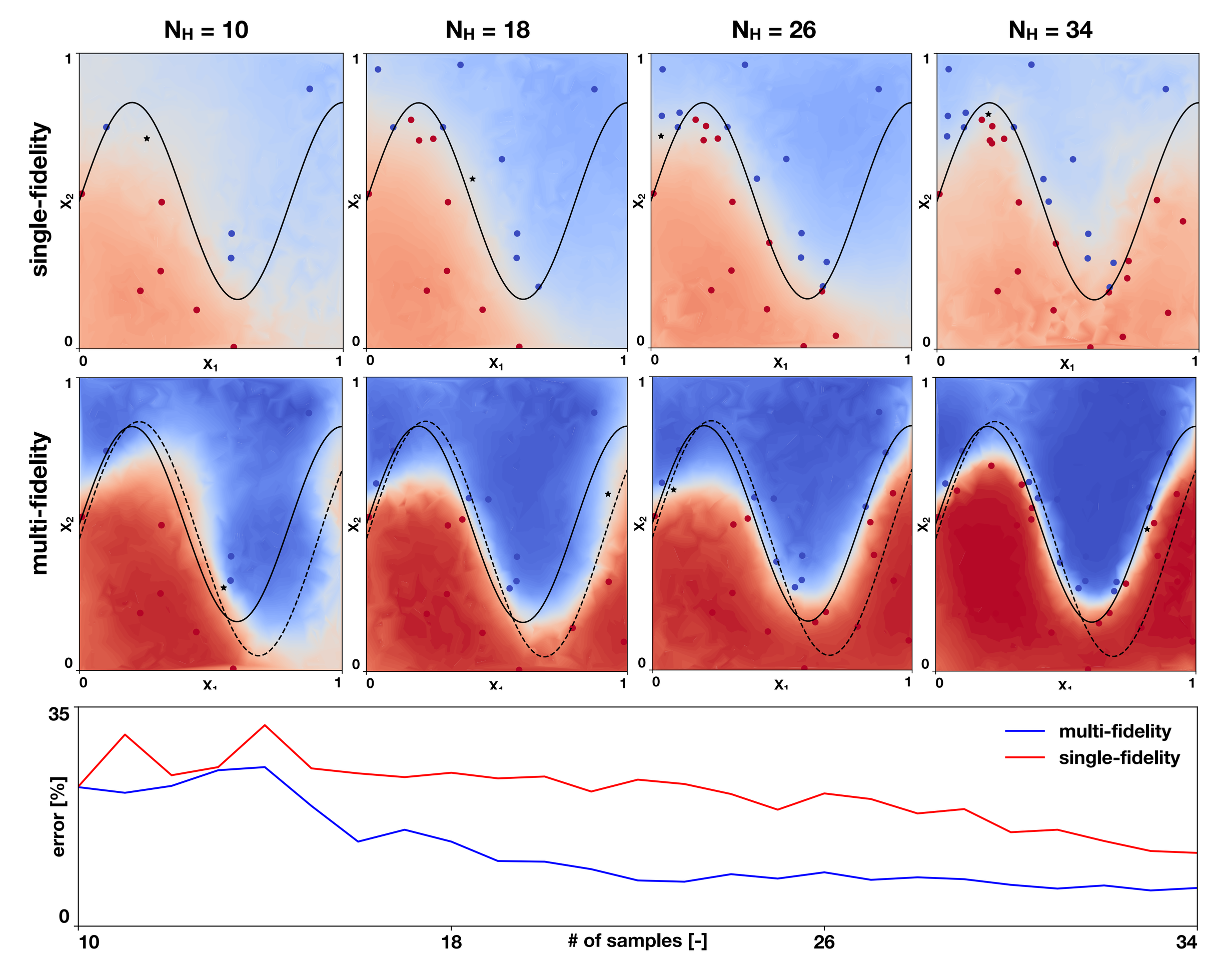

Figure 1 - Active learning of the synthetic example. We define an arbitrary boundary to test the performance of the multi-fidelity classifier. The high fidelity boundary is shown with a solid line and the low fidelity boundary is shown with a dashed line. For all steps the multi-fidelity classier presents a sharper boundary (middle row) and is more accurate (bottom row) than the single-fidelity classifier.

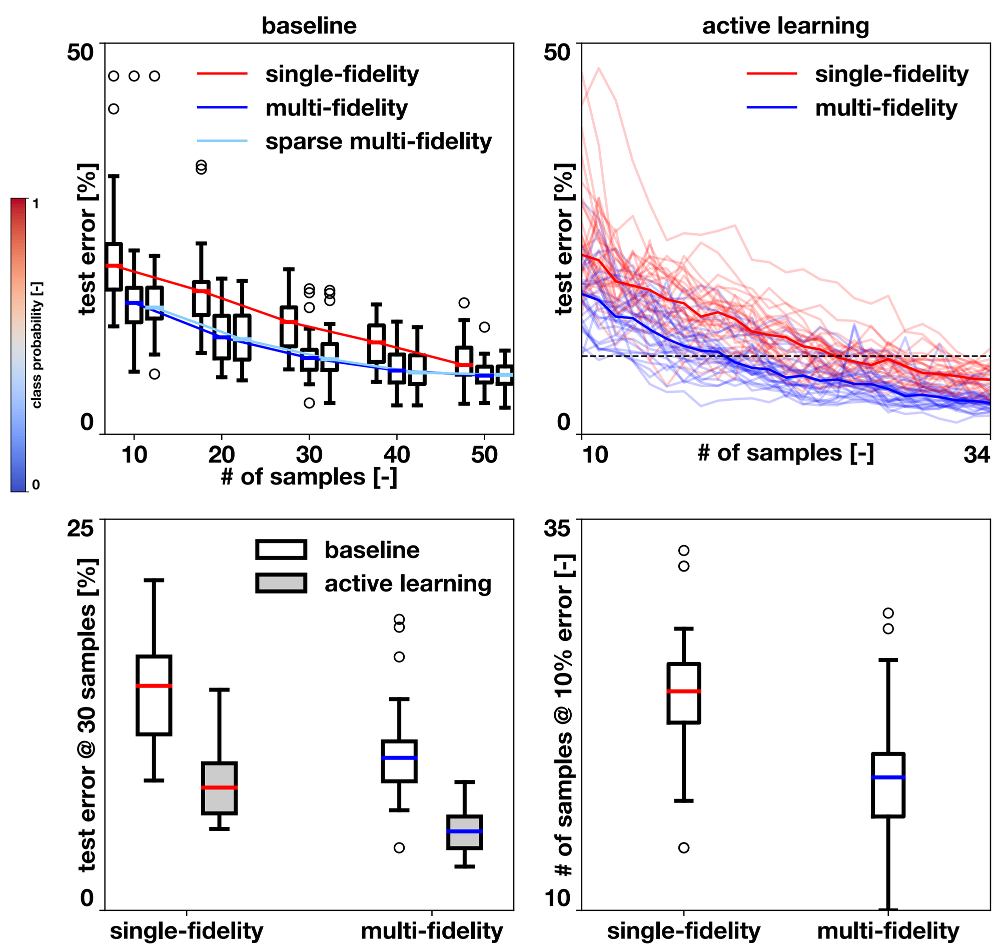

Figure 2 - Accuracy of the synthetic example. The top left shows box plots for different numbers of high-fidelity samples with no active learning. The multi-fidelity classifiers always outperform the single-fidelity classifier. The top right panel shows 30 active learning trajectories for each classifier type. Both classifiers reduce their error when combined with the active learning strategy. We compare the accuracy of the classifier trained with 30 samples with and without active learning. The active learning approach achieves significantly higher accuracy. The bottom right panel quantifies the difference in accuracy between the single- and multi-fidelity classifiers by counting the number of samples required to achieve 10% error when using active learning.

In an application to cardiac electrophysiology, the multi-fidelity classifier achieves an F1 score, the harmonic mean of precision and recall, of 99.6% compared to 74.1% of a single-fidelity classifier when both are trained with 50 samples. In general, our results show that the multi-fidelity classifiers outperform their single-fidelity counterpart in terms of accuracy in all cases.

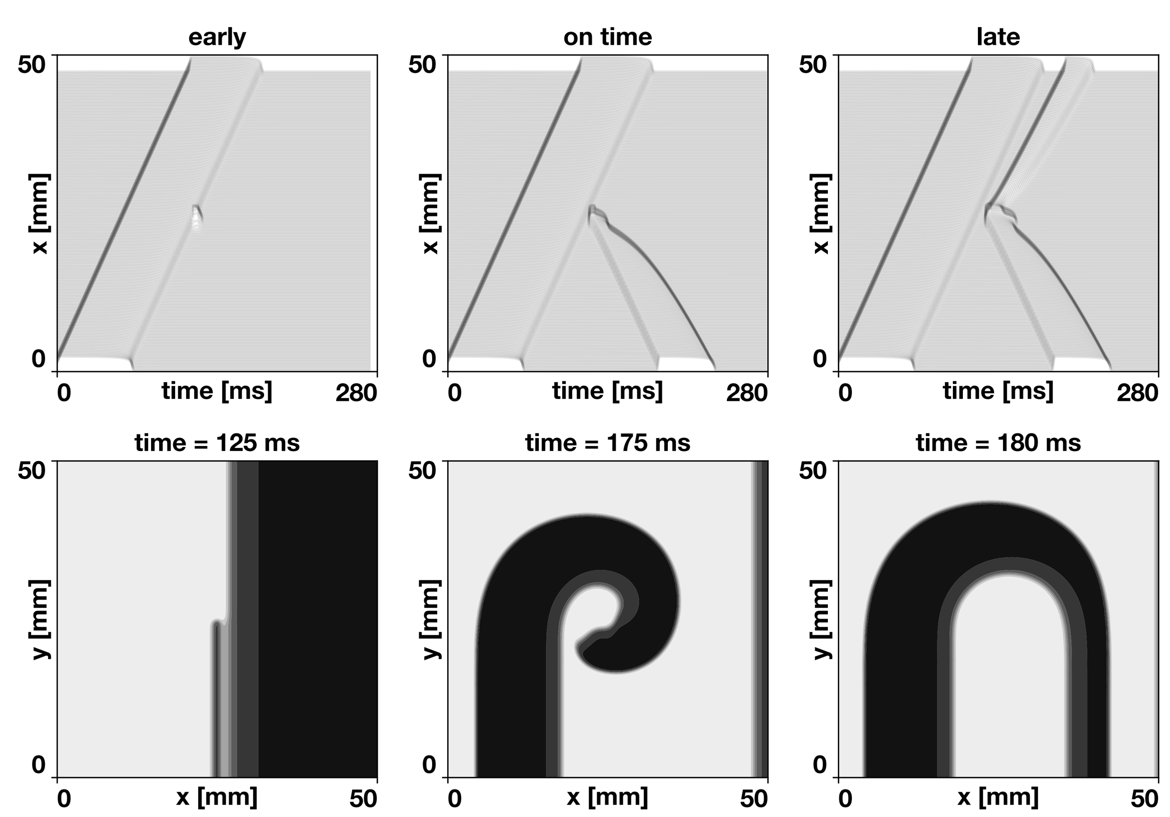

Figure 3 - Low and high fidelity models of the cardiac electrophysiology example. The top row illustrates the low fidelity model, a one-dimensional cable, where the vertical axis represents the position on the cable, horizontal axis represents the time elapsed, and the lines represent the action potential. The bottom row shows the high fidelity model, a two-dimensional patch of tissue, where the contour plots represent the action potential at a given time, with black being activated and white representing the resting state. Panels on left shows a secondary stimulus that is applied too early to generate a spiral wave. Middle panels show a case when the stimulus is applied within the vulnerability window and a spiral wave is created. Right panels shows a stimulus that is applied too late, failing to create a spiral wave.

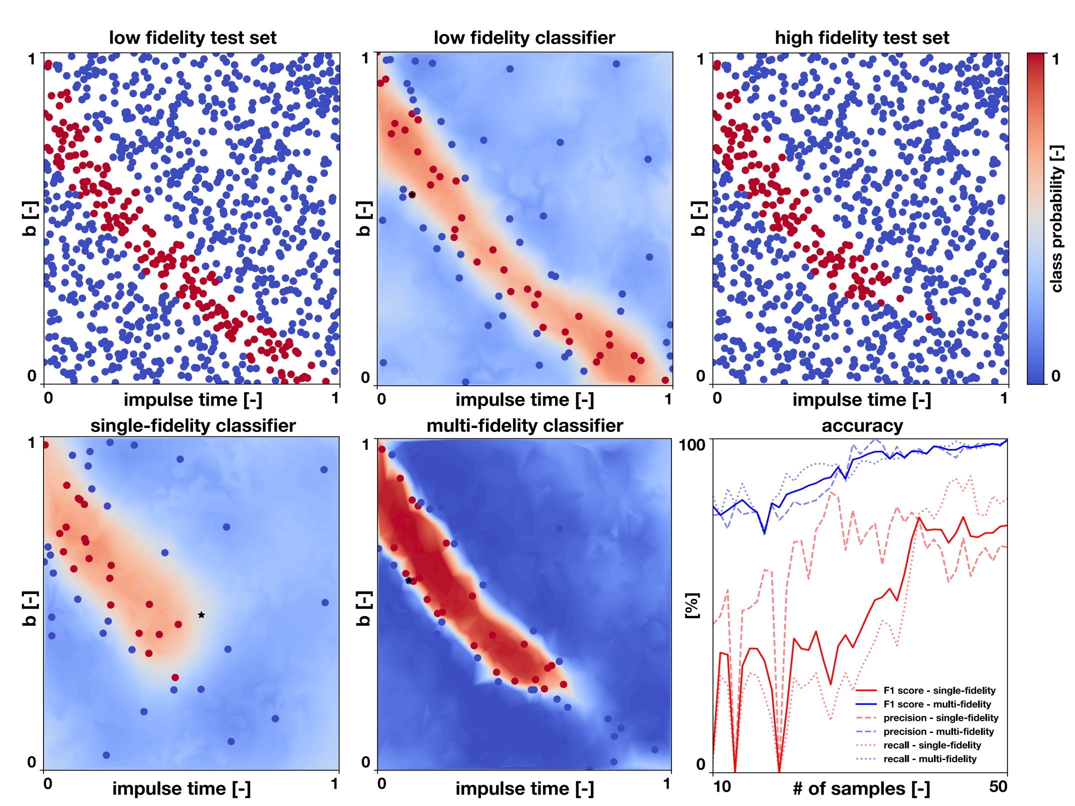

Figure 4 - Classifiers of the cardiac electrophysiology example. Top left, test set used to evaluate the low fidelity classifier, with N = 1000. Top middle, single, low fidelity classifier trained with active learning to obtain N_L = 84 used in the multi-fidelity classifier. Top right, test set for the high-fidelity data. Middle left, resulting single-fidelity classifier with N_H = 50. Middle panel, resulting multi-fidelity classifier with N_H = 50. Middle right, sparse multi-fidelity classifier trained with N_H = 50. Bottom row, accuracy comparison between single-, multi-fidelity and sparse multi-fidelity classifiers, showing precision, recall, and F1 score.

Discussion

We present a novel multi-fidelity classifier using Gaussian process priors. While the multi-fidelity paradigm has been proposed for regression and uncertainty quantification [1], this work represents one of the first attempts to formulate and implement a fully-Bayesian multi-fidelity classifier. Based on an autoregressive Gaussian process prior that enables the seamless integration of data with different levels of fidelity, our classifier outperforms the single-fidelity classifier both with synthetic data and in an application of cardiac electrophysiology. To strike a balance between classification accuracy, data efficiency, and computational cost, we propose an active learning approach that takes advantage of the predictive uncertainty in our predictions [3]. This approach significantly reduced the error in both examples, showing a good combination of exploration of the parameter space and exploitation of the boundary. Although the multi-fidelity and sparse multi-fidelity classifier work better in all examples presented here, the differences are greater in the cardiac electrophysiology example. This suggests that multi-fidelity classifiers are advantageous when there is class imbalance: only a small region of the parameter space is labelled with a particular class. We envision that this new tool will enable researchers to study classification problems that would otherwise be prohibitively expensive. Source code is available at https://github.com/fsahli/MFclass.

References

[1] Kennedy, M. C., O'Hagan, A., 2000. Predicting the output from a complex computer code when fast approximations are available. Biometrika 87 (1), 1–13.
[2] Hoffman, M. D., Gelman, A., 2014. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. Journal of Machine Learning Research 15 (1), 1593–1623
[3] Kapoor, A., Grauman, K., Urtasun, R., Darrell, T., 2007. Active learning with gaussian processes for object categorization. In: 2007 IEEE 11th International Conference on Computer Vision. IEEE, pp. 1–8.
[4] Snelson, E., Ghahramani, Z., 2006. Sparse gaussian processes using pseudo-inputs. In: Advances in neural information processing systems. pp. 1257–1264.