Surrogates:

- Identified basic classification of surrogates:

1- Classification based on data selection method (automated, human intervention):

Most surrogates models require human intervention for the selection and adjustment of training data such as in [8]. However, an active approach to surrogates learning that enables automatic training data selection has emerged in GP surrogates as in [4], [5]*.

2- Classification based on the modeling approach:

For example: Neural network surrogate [8], Gaussian process surrogate [4], support vector machine surrogate [7]. Multiple other categories exist.

3- Classification based on architecture (Fixed architecture vs variable architecture).

Most surrogates used in scientific workflow use a fixed architecture such as [1]-[4]. However, a recent paper [6] successfully used NAS (Neural Architecture Search) in order to infer the architecture of the neural surrogate.

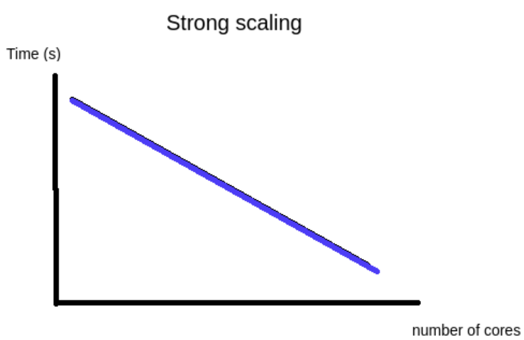
Previous RADICAL work [9] considered over 100 existing and well understood surrogates and classified them based on the algorithmic area - for example: partial differential equations, particle dynamics… etc. In addition, this paper classified the surrogates into three action-based categories: (1)Improving simulation with Configurations and (2)Integration of Data, Learn Structure, Theory and Model for Simulation, and (3)Learning to make Surrogates. The categories cover eight patterns for the link of ML to the simulations: (1.1)MLAutoTuningHPC: Learn Model from Data, (1.2)MLAutoTuningHPC: Learn Configuration, (1.3)MLAutoTuningHPC: Learn Model from Data, (2.1)MLAutoTuningHPC: Smart Ensembles, (2.2) MLaroundHPC: Learn Model Details, (2.3)MLaroundHPC: Improve Model or Theory, (3.1) MLaroundHPC: Learn Output from Input (parameters)and (3.2)MLaroundHPC: Learn Output from Input(fields)

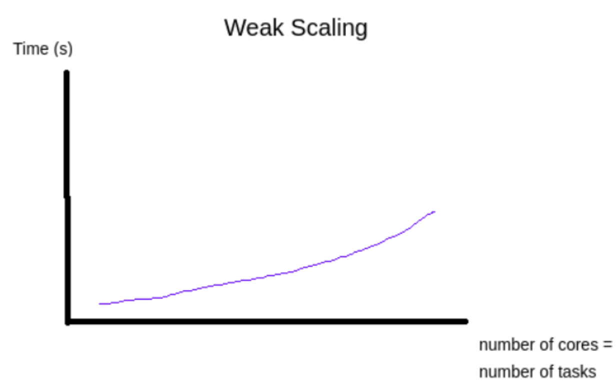
DeepdriveMD:

- Thoroughly read “DeepDriveMD: Deep-Learning Driven Adaptive Molecular Simulations for Protein Folding”
- Started reading the project code

Notes:

1- I’m unable to obtain the graph for strong and weak scaling. I will update this report with the graph once the I’m able to use EnTK again. Please see the two graphs below that depict my expected results.





2- I could benefit from your feedback regarding the surrogate classification above. Is it specific enough? Should it (eventually) be as elaborate as in [9]?

Resources:

[1] Machine learning and data science in soft materials engineering

[2] A high-bias, low-variance introduction to Machine Learning for physicists

[3] Combining Machine Learning and Physics to Understand Glassy Systems

[4] Quantifying the uncertainty in model parameters using gaussian processbased markov chain monte carlo in cardiac electrophysiolog

[5] Multi-fidelity classification using gaussian processes: accelerating the prediction of large-scale computational models.

[6]Up to two billion times acceleration of scientific simulations with deep neural architecture search

[7] Support vector machines for surrogate modeling of electronic circuits

[8] Artificial Neural Network based surrogate modelling for multi- objective optimisation of geological CO_2 storage operations

[9] Learning everywhere: A taxonomy for the integration of machine learning and simulations

[10] https://github.com/braceal/DeepDriveMD/tree/master/deepdrive