ME 8813–Machine Learning Fundamentals for Mechanical Engineering
Homework Assignment 2
(Due 2/26/2024, Monday, 11:59pm)

Your goal in this homework is to build several different models that can predict the stiffness values using the dataset provided. The dataset, given in file "HW2Dataset.csv", contains 8900 data points, each being a 15-element vector reflecting the salient measures of material structure (8900x15 array), and their corresponding scalar finite element-predicted elastic stiffness values (8900x1 array).

The 15 regressors are expected to generally show increasing levels of importance in predicting the output that reflect salient measures of material structure. You are not required to use all the regressors in your models. In building your models, please employ suitable train-test splits and cross-validations to demonstrate your model performance.

Specifically, please demonstrate the following strategies:

1) Use simple linear regression methods with or without regularization. Explain the rationale behind your choices.

2) Use the Gaussian Process regression method. Explain your choice of the kernel and comment on the methods used for the optimization of the hyper-parameters in the kernels. Hint: GPR models might face convergence issues when dealing with large datasets. You might want to select an optimal training dataset first that sparsely populates the input feature space. In other words, a 'de-clustered' dataset may improve the robustness of the GPR model. The most popular approaches to de-clustering a dataset include the k-Nearest Neighbors and k-means algorithms, both of which are available in the python package sklearn.

3) Use Neural Networks. Discuss the choices made in your network architecture (number of layers, activation functions, etc.).

Discuss the relative pros and cons of all the models you have built for this homework. Which ML method performed the best? Why?

Submission guideline: Submit your results in a single pdf with the appropriate figures and discussion. You will also need to submit a separate python code used for the implementation of the different models. You may use a Jupyter notebook file with different cells indicating different questions. You may add comments and titles to indicate which questions/case studies you are studying/discussing. The python code needs to run without any bugs.