

CH5650: Molecular Data Science and Informatics
Assignment#2
Maximum Marks – 15

Timeline: March 12, 2023

Data: A water cluster of ~1800 optimized structures is provided. All of them are given in the POSCAR format.

Problem Statement

Build machine learning models using the given data to predict the energy of a water cluster and forces on each atom of the clusters. The model should be trained based on energy only. Use 80% data to build the model and test the model performance for the remaining 20% data. Write a detailed description of your method, fingerprinting and model building. You may use the Behler-Parrinello symmetry function for fingerprinting. Define a loss function or cost function for model building and draw the function during the training of models. Report the parity plots for training and test data sets.