Abstract

Computing the conformational free energy surface of a molecule in solution has posed a challenging problem in physical chemistry. The goal of our work is to assimilate discrete simulation data into a compact model which can be used for further analysis. In this thesis, support vector machines (SVMS) were used to compute the FES of several data sets representing discrete free energies of increasingly complex polypeptides in solution. The results of the function generated were found to depend on the number of training points, number of dimensions, the kernel function, periodic boundary conditions, cross validation fold number and the parameter grid space. This thesis provides an overview of the number of folds needed for cross validation, popular kernel functions, and example parameter grids which are important for creating the best model. An in house convex solver utilizing the PULP library was created to provide the user more customizability compared to sklearn (an open source library that provides machine learning models) when selecting the kernel function, objective function, and periodic boundary conditions. The program also allows gives the user various visualization tools to help view the PD nature of the kernel, alpha distributions, dimensional spread and prediction vs. test output.