CHEM 4542/6109 Computational Chemistry

Take-home project

Machine Learning for Predicting Organic Photovoltaic Efficiency

Due: Dec 17, 2024, Tuesday

Introduction:

In this project, you will train the data for predicting organic photovoltaic efficiency. The data contains the following features for 515 data points (Tables S3 and S4 in Supporting Information) from a literature reference (https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202213953).

E_g: the photovoltaic optical gap, referring to the onset of the experimental absorption spectrum.

ΔE_{DA}: the smaller value of DFT HOMO-LUMO energy differences between the donor and acceptor.

ΔE_{ST}: the DFT energy difference between lowest singlet (S_1) and triplet (T_1) states.

V_{OC}: open-circuit voltage (photovoltaic parameter)

J_{SC}: short-circuit voltage (photovoltaic parameter)

FF: fill factor (photovoltaic parameter)

You can also supply more feature inputs (e.g., fingerprint) beyond the given ones to boost the training. The target data is labelled by the power conversion efficiency (OPE). The scikit-learn source codes for different ML regression algorithms (linear regression, K-nearest neighbours, regression tree) along with cross-validation search are provided. The DFT optimized atomic coordinates of donor and acceptor molecules are also given.

This project is to explore the data and come up with your own **regression** or **neural network model** with best parameters (including model parameters, training hyperparameters and feature sets, not limited to the feature provided in Supporting Information) for predicting OPE.

Project requirement:

1. The final project performance will be evaluated by assessing your written codes (40%) and project presentation (60%). They will be evaluated according to

- presentation clarity and coherence, thoughtness of the idea and the data analysis
- effective use of computational concepts and tools
- clear demonstration of your prototype code in solving the project problem
- lack of grammar and spelling errors.

2. Please submit **video link** (such as youtube) and **ML code**. We do not accept video files. We highly recommend that you record your presentation video on Zoom, where you can share your ppt slides and Jupyter page.

3. The project presentation video shall be no less than 10 min and no longer than 15 min. The prerecorded video must show your face through the entire presentation. The project video without showing your face will be directly failed.

4. You shall present your project model design, computational details and results by using ppt slides as well as Jupyter notebook (or your own python platform). In the video, you must demonstrate executing your program.