

Computational Chemistry and Materials Modeling

Lab 5, due date is set in Canvas LMS

Topic: Materials data science

Notes: Upload solution as a single file “YourName.zip”. Provide absolute minimum of supporting info – no copies of work folders. Compare results with published experimental and theoretical data. Solution must be submitted as article-style report supplemented by required technical files: xyz- and cif-geometries, program run log- or out-files, extra figures etc. Be prepared to give a 5 min presentation of everything that you consider nontrivial in your work.

Take a computational dataset with at least 100 entries. Use 50-90% of data for training and the rest for validation. Select a scalar target property (e.g. water adsorption energy or optical gap) and determine the best vector descriptor:

- Guess a single elementary descriptor (primary feature, e.g. d-band center or HOMO-LUMO gap) and perform linear regression analysis (determine coefficients and their standard deviations, RMSE and MAE, R^2 -parameter).
- Add several other primary features and use SISO or other software to identify the best descriptor and the predictive model. Compare RMSE and MAE with the above simple model. Perform this for two different sets of mathematical operators (generators of complex features).
- Simplify the identified descriptors and provide their physical meaning. Compare RMSE and MAE with the above models.
- Define the “best” value of the target property (e.g. an intermediate water adsorption energy or maximum sunlight harvesting optical gap) and compare performance of the above models for prediction of the best material.

Sample solution: See Lab5_wateradsorp.zip