

**C142/242: Machine Learning and Optimization for Chemical Problems**  
**Final project guidelines & tips**  
**Assigned March 19 and Due April 25**

**Jupyter Notebook (all students):**

To receive full credits, your codes should be complete, organized and well-documented such that any person with reasonable python experience can understand without the need to reference to your report/presentation.

- Remove all debugging or unused lines in your codes, and section them by: data processing, model, training, output analysis/visualization if applicable. If you are a grad student doing 2 methods for your project, annotate the two clearly.
- Show your learning curve. You should have a reasonable learning curve if you have trained your model properly.
- **Please condense and highlight your training results.** This can include but not limited to average train/test errors, accuracy or any relevant evaluation metrics you have defined, comparison between prediction and the ground truth... Use appropriate figures and tables to present them. If paragraphs/markdowns are included, be concise.
- Make sure your final pdf does not cut off any of your codes and outputs.

**Powerpoint presentations (Grads):**

This is expected to be a 10 minute presentation to the class: 8 minutes presentation and 2 minutes of questions. It should be in the style of a graduate student research seminar. We have placed some examples from last year to get a sense of content.

**Create a Github (Ugrads/Grads): Extra Credit!** Git has largely become the community standard for version control in code development and dissemination to others.

**Final Project Report (Ugrads):**

Written reports should contain the following sections: introduction, methods, results, and discussion, and be a minimum of 3 pages (excluding references). Font 11-12pt, 1.5 spacing, 1" margins. References as many as you want.

- The introduction should describe the background information on the problem --- why study the problem and with the proposed method, what has been done (with proper literature citations) --- and the objectives of the current project.
- The method section should provide succinct but sufficient information on your data source and representation, model architecture and training scheme. Cite public packages you have used. List the hyperparameters for training.
  - To receive full credits, you should train with data of at least 4 heavy atoms.
  - Your notebook implementation should be consistent with what you describe here.
- Tabulation and/or plots of data/observations/predictions and summary of findings.
  - Report your results in the correct format (RMSE and consider energy unit).
  - Aim for quality training results (error  $\leq 5$  kcal/mol).
  - Your notebook should show any conversion/calculation required to get to the reported training results.

- Discussion can contain higher-level analysis and interpretation of the results not already stated in your result section. For example, to what extent your results achieve the objective or resolve the problem.