

Daily Log

Monday, February 10

Started removing the parse errors that I ended with last week.
Removed offending column. Unfortunately, the dataset I found still does not seem to match up with the intended test dataset.
Downloaded NCI dataset as alternative.

Tuesday, February 11

Debugged SMILES string parser.
Spent some time completing the DeepChem tutorials to understand their dataset object fields.
Briefly investigated how to roll back version of dependencies in Anaconda.
Reshaped some labels but not their features.
Changed some function calls so the code would run in Python3.

Thursday, February 13

Grappled with casting bugs.
Rolled back version of Python to 2.7 to match the code's author.
Continued changing function calls so the code would run in Python3.

Timeline

Date	Goal	Met
Feb 3	Analyze benefits provided by multi-task learning and begin hyperparameter tuning.	Compared evaluation losses; did not start hyperparameter search.
Feb 10	Identify and start implementing alternative neural methods.	Yes; installed DeepChem to investigate MPNNs.
Feb 17	Understand code from MPNN tutorial. Start building neural fingerprinting.	No; this goal has been expanded to fill the next 2 weeks.
Feb 24	Build first example MPNN.	
Mar 2	Implement molecular fingerprinting	

Reflection

I have spent this past week trying to work with an MPNN implementation that I found in the user-contributed section of the DeepChem library. Because it is user contributed, I don't have the same guarantee of quality as I would have for the rest of the library. In particular, I have been struggling to get the working tutorial example working because the dataset that they are requesting is found in a file that I don't have access to. I was able to find the dataset elsewhere and was able to reformat it to match the expected format of the code I'm trying to use, but regardless I'm getting obtuse error messages. It also took me an embarrassingly long amount of time to recognize that the code has been written in Python2, so I spent some class time installing different versions of packages in order to ensure code compatibility while simultaneously working to rewrite the code so it would run on my existing installation of Python3. (For example, it took me quite a bit of time to realize that Python2 uses eager evaluation on methods like `map`, so I had to add in additional casts to make the code run in Python3).

I am still in the process of creating a working first example, because I have not yet vanquished all the bugs. I want to spend the next couple of classes completing this task. If this turns out to be impossible, I will instead turn to building my own implementation using the example as shell code. Once I have a working MPNN, I will turn to the example from Duvenaud et al. to implement molecular fingerprinting.