Chemistry 213/313 Homework Assignment 2 Due: April 25, 2025

In this homework assignment, you will gain experience with basic deep learning methods. To submit the assignment, export each notebook to HTML and upload the resulting file to Canvas. Your code should be well-commented through Python comments and markdown cells.

1. Construct a single hidden layer neural network to approximate the two-dimensional trigonometric function, $f(x_1, x_2) = \sin(x_1) + \sin(x_2)$. Specifically, develop a Jupyter notebook that does the following.

   (a) Create synthetic data, $(\vec{x}, f(\vec{x}))$ by uniformly sampling $\vec{x}$ from $D = \{(x_1, x_2) | x_1 \in [-3, 3], x_2 \in [-3, 3]\}$.

   (b) Prepare a single-layer neural network with a mean squared error loss function by defining a subclass of the PyTorch nn.Module class; also prepare a stochastic gradient descent (SGD) optimizer. These should both be PyTorch objects.

   (c) Make an 80:20 train-test split and train the neural network. Plot the training and test losses at each epoch.

   (d) Experiment with different choices of activation function (e.g., sigmoid, ReLU, etc.).

   (e) Produce a contour plot of the function approximated by your best model, and compare this to the target data.

   (f) Try increasing the width and depth of neural network and see how it improves the performance by showing the training and test losses plot and the contour plot.

2. Construct a multi-layer perceptron (MLP) to predict solubility using the dataset discussed in Homework Assignment 1. To this end, develop a Jupyter notebook that does the following:

   (a) Gets the data, loads it into a pandas dataframe, and also saves it locally (or uses the data you saved for Homework Assignment 1).

   (b) Makes an 80:20 train:test split and normalizes the features appropriately.

   (c) Constructs an MLP using PyTorch to predict solubility. Train the model and plot the resulting predictions against the actual values. Comment on the performance of the model compared to the linear model you built in the first homework.

   (d) Experiment with different architectural choices (e.g., different network widths and depths, different activation function) and regularization techniques. Discuss how these changes affect the performance of the model.

3. Construct a convolutional neural network (CNN) to classify protein sequences as soluble or insoluble based on the dataset discussed in White Chapter 7. Develop a Jupyter notebook that does the following:

   (a) Downloads the dataset, preprocesses it (e.g., creates labels, shuffles the data, etc.), and saves the resulting dataset locally.

(b) Constructs a CNN to classify the protein sequences, trains the model, and evaluates its performance on a test set.

(c) Experiment with different architectures (e.g., different numbers of convolutional layers and filters) and regularization techniques introduced either in class or White Chapter 7. Discuss how these changes affect the performance of the model.