Final Project Challenges

Setup:

- Please download and install Anaconda: https://www.anaconda.com/products/individual
  Run Anaconda Navigator and then launch Jupyter Notebook (*install if needed*). This will open a page in your browser.
- Save the class Jupyter notebook to a folder. The Jupyter Notebook page will show you a directory list. Use the directory controls to navigate and open the class Jupyter notebook.

Submission:

- Please submit a Jupyter notebook with the code that you've written. Also, for any written answers to questions, please include as markdown text in the notebook and indicate which question you are answering (e.g. "1c"). If you want to work collaboratively you may open these as a google colab but you will not be able to control the robot from the colab (https://margaretmz.medium.com/running-jupyter-notebook-with-colab-f4a29a9c7156)
- For all challenges, save acquired data and graph images for submission. Additionally, take videos of Challenge 4 (Active Learning Loop) for your final presentation, and as proof of completion.

Challenges:

1) **Measurement Uncertainty**
   a. Write a script to create a sample with an equal amount of acid and base. Total volume should be 2 ml.
   b. Write a script to follow these operations in a loop with 10 measurements:
      i. Dip the pH sensor in the DI water well
      ii. Dip the pH sensor into the sample from (a), measure pH and record.
   c. Plot the results from (b). What is the mean and variance of these measurements? Please include this information as markdown text in the Jupyter notebook.

2) **The Henderson-Hasselbalch (HH) Equation**
   The HH equation is given as: $pH = pK - log\left(\frac{[Acid]}{[Base]}\right)$, where the output gives the pH of the sample, pK is a constant, and [Acid] and [Base] are the concentrations of acid and

base. For these experiments we will work with the percentage of acid in the sample to create samples varying in composition from a percentage of acid of 10% through 80%.

a. Write code that will create 8 samples of mixture amounts: acid [10%, 20%, … 80%]
b. After each sample is made, have the system dip the sensor in the DI water well, and then measure the pH of the sample. Plot the measured pH as a function of the [acid]/[base] ratio. *(Note that you must convert to concentrations to calculate this ratio)*
c. Run the code created during steps (a) and (b) (this should be one for loop accomplishing both tasks together). Plot the results.
d. We will now fit the HH equation to the data. Write a function that takes in the [acid]/[base] ratio **x** and returns the **pH**, following this equation:
    i. Func(x, pK) = $pK - (x),\ where\ x\ =\ [acid]/[base]$
e. Follow the example for the scipy function "curve_fit" to fit this function to the data you have measured: scipy.optimize.curve_fit — SciPy v1.7.1 Manual (use default method and no optional parameters.)
f. What is the pK value discovered?

3) Gaussian Process

a. Fit the data with Gaussian process regression using an RBF kernel. Optimize the hyperparameters and plot the results. Output the GP hyperparameters: kernel length scale, kernel variance, noise variance.
b. Do the same, but this time with the data for **x** corresponding to acid percentages of [10%, …, **60%**]
c. Bad assumptions:
    i. In (a) and (b) we optimized to find the best values for the hyperparameters. Let's see what happens when we pick a poor value. Here we again will use the partial data used in (b). Using the following lines of code, fix the kernel length scale to first a value of 0.1 and then 10. For each case, optimize the rest of the parameters and plot the results. What is the impact of fixing a small or large value for the length scale?
    m.rbf.lengthscale = 0.1
    m.rbf.lengthscale.fix()
    ii. Now let's see what happens when we make a poor assumption for the kernel. Again use the partial data from (b). Let's set the kernel to the standard periodic:
    StdPeriodic(1)
    This imposes the assumption that the pH function is periodic. Optimize all the hyperparameters and plot the result.

4) Active Learning: Optimizing

In this section we will combine the GP of (3a) with active learning.

a. We would like to find the value of **x** where pH = 4.75
    Define an acquisition function and discuss with the TA or a professor. For this challenge combine exploitation and exploration.
b. Combine the acquisition function with the GP of (3a). Hint: use the demo code from the active learning lecture. Allow active learning to select any acid

percentage in the range 0% - 90% acid). Have active learning make samples in sequence, following the GP and active learning loop. Define an exit criterion appropriate for this type of study (what defines success?) What result do you get?