*Instructions:* **Machine Learning for Physical Sciences, class project**

Contents of project folder:

Files:

* *Create\_CSDM.m* : file to preprocess timeseries data into normalized, vectorized cross spectral density matrices
* *DataSet01\_05.png*: graphic of the 5 data sets. training and test sets are along the same path at different times
* *Design\_labels.m*: file to create data labels corresponding to ship range indices
* *Fig\_DataSet01.jpg*: an example of DataSet01 results using a basic SVM program in Python, for four kernel functions
* *localization\_SVM.py*: a Python file to conduct localization using SVM
* *SVM\_demo.m*: a Matlab file to conduct SVM
* *Tracks\_SIO209.docx*: a Microsoft Word document with details about the ship tracks for each data set  
  Folders:
* *libsvm\_matlab*: the necessary functions to run Matlab SVM. Ensure that this folder and its content are in your Matlab path before running.
* *data:* contains data for conducting Machine Learning ship localization  
   Existing:
  + *AcoustData*\* are raw timeseries (received pressure, μPa) data files, one for training and one for test
  + *GPS\_Range*\* is for creating data labels   
    You must create:
  + *Mapping\_range\_labels.txt* generated by *Design\_labels.m*
  + *SBCEx16\** generated by *Create\_CSDM.m*. Use as input for your machine learning code.
  + *test\_labels*\* and *training\_labels*\* generated by *Design\_labels.*m. They are labels for the data. These will need to be converted to one-hot vectors for use with FNN.
  + *test\_Ranges.txt* and *training\_Ranges.txt* are generated by *Design\_labels.m*. They are the ranges of the ship.

Instructions:

***Make sure to set the data directory to the correct location in every file!!***

1. Run "Create\_CSDM.m". Run this code *twice*: first, set TRAIN= true and set the filename to your training set (AcousticData\_255190605\_255192105).
2. Second, set TRAIN= false and set the filename to your test set (AcousticData\_256093210\_256094810). This code will preprocess the raw data by our method. You can change the selected frequencies, the signal duration, and the number of snapshots. \*If you change the number of snapshots, also change it on your label data.   
   There are many methods to preprocess data, so you may play with other methods if you choose in addition to this one. However, running this file will create files with time references, and step 2 will not run without those (unless you modify it...).
3. Run "Design\_labels.m" to generate the data labels. You can choose the label resolution (or use the raw range data in regression model). You can set the number of average data snapshots to match your data preprocessing.
4. Run "SVM\_demo.m" or "localization\_SVM.py" to try classifying the acoustic data into range classes.
5. Play with parameters or try a new machine learning algorithm! See if you can improve the localization success, try a completely new dataset, or test a different machine learning method.