Michigan State University **Feb 23, 2024**
CMSE381 - Data Science

CMSE381 - Quiz 5

I will adhere to the Spartan Code of Honor in completing this assignment.

Signed: _____

1. 3pts. I wrote some code to setup k-fold cross validation for her learning model, but she made a mistake in my code. What's the error?

```
def mykfold(n,k):
    # Input: integers n and k.
    #        This version is only going to allow us to work with
    #        a $k$ that is actually divisible by $n$
    # Output: a list of the train/test splits to be used.

    # Make an array of the indices:
    all_my_indices = np.array(range(n))

    # Length of my folds:
    L = n // k

    AllSplits = []
    for i in range(k):

        # Extract the test set
        test_set = all_my_indices[i*L:(i+1)*L]

        # Get the training set by deleting the points used for the test set
        training_set = np.delete(all_my_indices, range(i*L,(i+1)*L))
        AllSplits.append({'train': training_set, 'test':test_set})

    return AllSplits
```

Didn't shuffle the training set before splitting

(a) 4pts. For each of the following properties, is it something describing k-fold CV (assuming $k = 5$ or $k = 10$), bootstrapping, or neither?

i. Resamples data with replacement

k-fold CV boostrapping both neither

ii. Used for measuring the performance of a model.

k-fold CV boostrapping both neither

iii. Used for creating an empirical distribution of a statistic.

k-fold CV boostrapping both neither

iv. Requires the ability to resimulate your data.

k-fold CV boostrapping both neither

v. Involves no randomness, so will provide the same result every time if given the same data set.

k-fold CV boostrapping both neither

2. 3pts. We're a third of the way through the semester and we've had our first test.

   - How do you feel like the class is going so far, either in general or for you specifically?

   - What is one thing you can do to be more successful?

   - Is there additional support we can provide to help with this?