**ECE 8527: Introduction to
Machine Learning and Pattern Recognition**

# HW No. 11: Clustering

For set #8, build three classifiers:

1. **Supervised K-MEANS**: cluster the training data for each class into N clusters (a total of 3N clusters).
2. **Unsupervised K-MEANS**: cluster the training data into 3N clusters using K-Means. Assign a label for each cluster based on a majority vote of the tokens in each cluster.
3. **Agglomerative Clustering**: Cluster the data in a bottom up fashion until you reach a total of 3N clusters. Assign a label for each cluster based on a majority vote of the tokens in each cluster.

Classify the /train, /dev, and /eval data using these three classifiers. Plot performance on the eval set for the range $N=[1,100]$. Identify the optimal value of N for each algorithm. Justify your choice of N for each algorithm.

Create a table that compares performance to CD-PCA, KNN and RNF on the same data. Comment on the computational requirements for classification (not training) for each algorithm. Compare the computational complexity to the three baseline algorithms.