<u>**Active Learning for Computer Vision Curriculum**</u>

<u>**Project 6: Online Active Learning**</u>

Total Points Possible: 50

<u>Datasets</u>: The datasets required for this project are included in the folder. For the feature matrix in each dataset, each row denotes a sample and each column denotes a feature.

<u>Problem 1 (10 points)</u>

What is online learning? How does online active learning differ from pool-based active learning?

<u>Problem 2 (40 points)</u>

The Multiple Biometric Grand Challenge (MBGC) is a popular dataset for face recognition under unconstrained natural conditions. We will use 25 randomly selected subjects from this dataset for our experiment.

Implement the Label-Efficient b-Sampling and the Logistic-Margin Sampling algorithms for online active learning. Test their performance on the MBGC dataset and use uniform sampling as a comparison baseline. Take the query budget as 1000. Repeat the experiment 3 times with different initial training, unlabeled and test sets and plot the average training size vs. accuracy graph.