Active Learning for Computer Vision Curriculum

Project 3: Pool-based Active Learning – Part 1

Total Points Possible: 50

Datasets: 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. Each dataset contains an initial labeled training set, an unlabeled set and a test set. Also, each experiment needs to be run 3 times (with different initializations of training, unlabeled and test sets) and the average results should be reported.

Problem 1 (50 points)

Implement the Uncertainty Sampling based Active Learning method by Holub *et al.* Run the algorithm for 1000 iterations and query one sample in each iteration. Use random sampling as a comparison baseline, where in each iteration, an unlabeled sample is selected at random. Plot the accuracy on test set vs. size of the training set graph. Test the performance of these algorithms on the following two datasets:

i) The VidTIMIT dataset is widely used for facial image recognition and contains images of subjects reciting short sentences under unconstrained natural conditions. We will use images of 25 randomly selected subjects in this experiment.

ii) The MindReading dataset is widely used for facial expression recognition and contains images belonging to 6 different emotion classes.