Active Learning for Computer Vision Curriculum

Project 8: Active Learning for Regression

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.

Problem 1 (50 points)

The Flickr dataset contains a collection of images from various categories together with corresponding popularity scores. Since the popularity is a continuous variable, predicting the popularity of a given image can be considered a regression problem.

Implement the following regression-based active learning algorithms on this dataset. Please refer to the paper by Cai et al. (discussed in the slides for reference on all the algorithms)

i) The algorithm based on maximizing the expected model change

ii) The Greedy algorithm where the goal is to select the new sample which has the largest minimum distance from the labeled set in the feature space

iii) Random Sampling, where unlabeled samples are selected at random.

Run the algorithms for 1000 iterations. Use the mean squared error (MSE) and mean absolute error (MAE) as performance metrics and plot the error graphs with training set size.