

Active Learning for Computer Vision Curriculum

Project 11: Generalizations of Active Learning

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)

Implement the QBC based active matrix completion algorithm. Use the k-NN, EM and SVD based methods as members of the committee. Use the SVD based completion algorithm to complete the matrix entries after each batch query. Use Random Sampling as a baseline where the entries to be queried are selected at random. Test the performance of the algorithm on the following two datasets:

- i) The Lena Image is a matrix of grayscale values. Randomly delete 60% of the entries in the matrix and treat them as missing entries. Use active matrix completion with batch size 50 and 50 iterations to reconstruct the matrix. Repeat the process for 5 random runs. Plot the reconstruction error vs. iterations graph. The reconstruction error is computed as the Frobenius norm of the difference between the ground truth and reconstructed matrices.
- ii) The “sparseRatingsMatrix” contains ratings given to movies by a set of users. This matrix inherently contains a lot of missing entries (denoted by 0 in the data). Since we do not know the ground truth values of these entries, restrict your experiment to only the known values in the matrix. Randomly delete 50% of the observed entries in the matrix. Take the batch size as 50 and repeat the process for 100 iterations. Plot the reconstruction error vs. iterations graph.