This code corresponds to the algorithm in the paper Jacobs\_Merkurjev in this zip file. Most of the code was written by Matt Jacobs.

This code uses the MNIST data set, but can be easily tailored to any data set.

To compile the code, use the command sh compileCommand\_mnist in the terminal.

To run the code, use the command sh compileCommand\_mnist2 in the terminal. This will run all the experiments and display the accuracy for each experiment, and then the average accuracy. Each experiment considers a different labeled set of the same size.

compileCommand\_mnist2 has several parameters:

1. MODE: consisting of three letters: (usually dvn)

First letter: either k or d

 k - convolution with symmetric weight matrix

 d - convolution with symmetric weight matrix squared

Second letter: either r, p or v

 r - Initializes the non-fixed labels with a random label.

 p - Initializes labels with the MBO algorithm

 v - Initializes the non-fixed labels by creating a voronoi diagram with the fixed labels as the seed points.  Every point is assigned the label of the fidelity point in its voronoi cell

  Third letter: either f, w or n

 f - reweight fidelity nodes assuming that number of points per class is known

 w - reweight fidelity nodes assuming that number of points per class is unknown

 n - fidelity nodes are not reweighted

1. Number of nearest neighbors to use (usually relatively small like 20)
2. Percent of elements with fixed correct labels (small < 1 percent)
3. max number of iterations to run mbo ( >100)
4. number of experiments to run (as many as possible)
5. algorithm ends if energy change is below this parameter (<=0.0001)
6. Min epsilon parameter for the auction algorithm (small, ex) 0.0000001 )

Additional Notes:

1. The input into the algorithm is a file which contains the K nearest neighbors information about each element, and the distance to them.
2. Information about how the nearest neighbors are computed is contained in the Nearest Neighbors Computation folder, just for reference. Numerical feature vectors for each data element are used as input to the file that computes nearest neighbors. (vl\_neighbor\_data\_mnist)
3. Accuracy is computed just by comparing the output to the ground truth (we have the correct labels for all the data points).