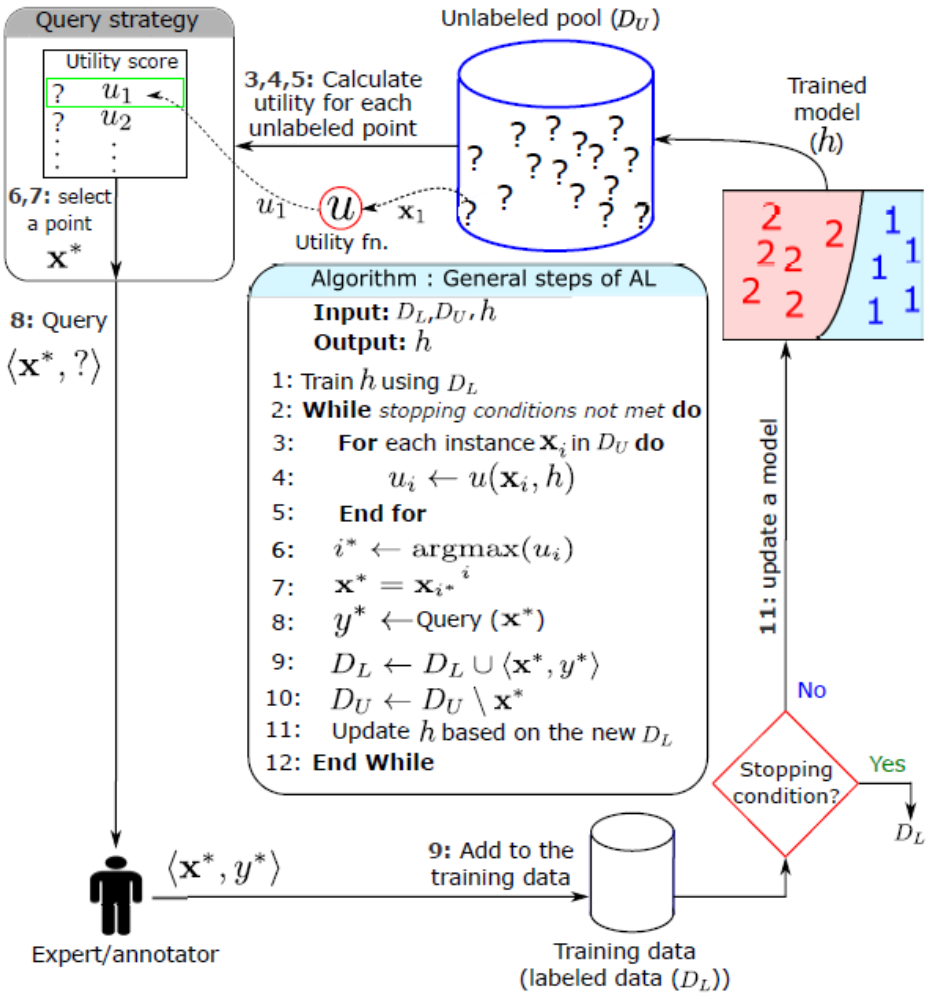
Assignment 8

How many labels do we really need, anyway? - **Active Learning**



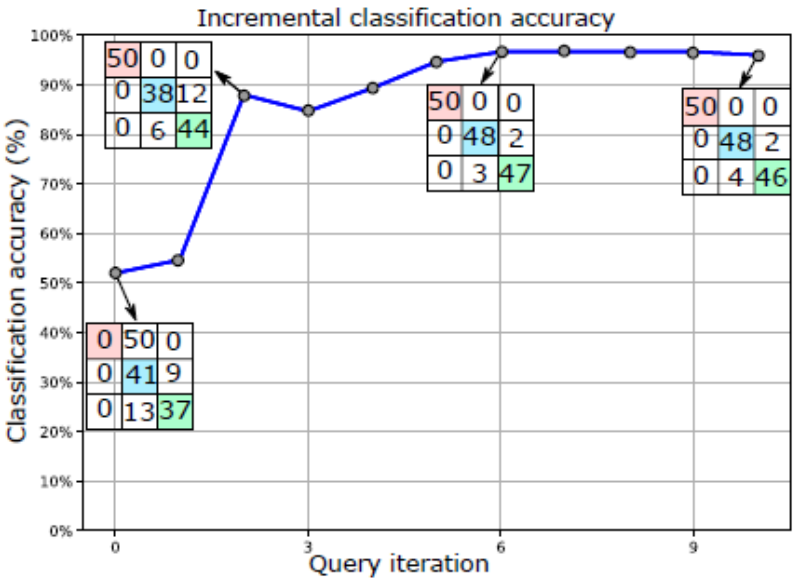
In this assignment you write code “from scratch” to implement an an active learning algorithm, and test it on the original MNIST dataset.

Specifically, you will:

- Get the MNIST dataset (the original, balanced – not MNIST-C).
- Write your own version of an active learning classifier. This should follow the basic outline shown in the figure above. Assume that the algorithm will start by querying 1 point at random. You may choose whatever utility function and classifier model you like. You do NOT need to write the actual classifier model itself from scratch.

- Run your code on the dataset and determine accuracy and a confusion matrix.
- Do this sequentially over a number of iterations sufficiently large to see performance flatten out, and plot accuracy vs. number of iterations. Show the confusion matrices at some selected points along the way.

This should result in a figure looking _roughly like the following:



(e)

You will turn in both the code and a short report, discussing what you've implemented, how well it worked, what you've learned, etc.

Remember, you must understand what you turn in – you may be asked to explain it to me and/or the class.