Problem Statement:
I want to use a classifying algorithm to be able to correctly identify handwritten math symbols on paper. If that goes well then I may expand into handwritten math expressions.

Data Preprocessing:
This is the link to my dataset: https://zenodo.org/record/50022#.XjzQG2hKhPa
The dataset is a collection of handwritten math symbols but they are online. I need to figure out a way to convert the online symbols to offline pixel images. Once I can do that for the entire dataset I will then move on and remove noise and resize the images if needed. I have downloaded the data and unzipped the tar file. Now I am working on converting stroke data into images. I successfully converted the stroke data into images. But my images are very large because the data has handwriting in different places. For example one person wrote ‘A’ in the upper left corner and another wrote ‘A’ in the lower right corner. I haven’t been able to figure out how to resize the data yet.

Machine Learning Model:
I will still be using a CNN. We are learning about it in class right now so I don’t know much about it yet. I am still in the process of working with my data so I haven’t been able to run my algorithm yet.

Preliminary Results:
Due to my training set containing 150,000 data points and each of which containing a very large image (900x900) I haven’t been able to run my model. It took a while to figure out how to convert stroke data into real images as pixel arrays. Once I figure out how to lower the size of each of my images I can start running my CNN.

Next Steps:
1. I need to resize the images to a low pixel count so that I can copy all my images into an X array of inputs without using all my RAM.
2. Once I do that I will run a preliminary CNN and look at my results.
3. Then I will work on adjusting the data and the model to increase my results.
4. Some possibilities include rotating the images so that the algorithm learn slight deviations from the vertical.