Thomas Coady, Ilyes Djebbar, Enyi Hou

MAIS 202 - Project Deliverable 1

1. Choice of dataset

We will be using the same Google Creative Lab dataset used to train the model for Google Quickdraw. This was the natural option, as it is easily accessible, and contains 50 million examples to train/test/validate our model. Downloading the dataset can be done in different ways, one of which is to install the GCloud SDK and then running a simple command to download the dataset, which comes out to about 24 gigabytes of information.

2. Methodology

a. Data preprocessing: Data may have to be preprocessed to account for different window sizes of users when drawing the doodles. It is unlikely that we will need to normalize the data, since either a pixel is filled or it isn't, there are no extreme values that would interfere with the neural network

b. Machine learning model: For our project, we will most likely use a Convolutional Neural Network (CNN) to train on the data. The data is stored as .json files with the drawing strokes with respect to time, so we will most likely just process the data to create images that represent the full drawing by the time the user is done and then feed that into the CNN. Then when we make the website and the user is actually drawing, we will periodically feed what the user has drawn so far into the CNN and have it predict the category based on that.

c. Evaluation metric: Since our model is a prediction model, and each drawing from our dataset is labeled, we can evaluate a model's accuracy based on the times it successfully identified the drawing. This is our primary approach, but we understand that our knowledge in this area might be limited, so we are open to learn and to incorporate other relevant metrics as we progress through this project.

3. Application

Once our model is trained, we hope to host it on a website and run it in the same way as Google Quickdraw. The user will be prompted to draw one of the 345 available categories using their mouse/trackpad and will have a time limit to do so. This can be turned into a game with points awarded for a drawing that is recognized by the model, with extra points for how quickly it gets recognized. If time allows we might even make it multiplayer, such that you are “racing” against another person to get your drawing recognized the fastest.