

Machine Earning  
Aaron Sager  
Connor Dupuis  
Brian Ho  
Max Gazeroglu

### Lab 1 Problem 3

The dataset is a collection of American Sign Language hand signs for letters “a” through “i”. The data is relatively available as the model designers can collect training data from their own hand signs. The amount of data available is linearly dependent with time, as the more pictures one takes, the more data one collects. The ground truth is generated by the designers rather than found from some available database. To improve the accuracy of the model, we will artificially increase the size of the data set by 10 times by taking each of the images, and rotating them between -5 and 5 degrees.

The error metric used will be accuracy, as the competition’s metric for success is also accuracy. The training data provided to us will be further split into training/validation and test data. This test data set will never be used by the model until after it is fully trained, at which point it will be used to determine how well the model performs. The results of this prediction on the test data will likely correlate well with the true test data that has been withheld from us. Moreover, we will employ cross validation to improve the accuracy of the model. This will involve splitting the training data into 5 buckets where 4 buckets will be used for training and the fifth for validation, repeated five times to decrease overfitting. If there is significant variation among data points in the provided training set, we may employ filters such as grayscaling to reduce variation.