

Homework 2

Kelsey Hawkins

Introduction

ASL is a prominent language in the American deaf community. With this in mind, I have a dataset of images containing the ASL dictionary, excluding the letters that need motion. The dataset was already flattened into greyscale values, ready for model input with a few preprocessing changes. This model will help translate ASL images to text for those who do not understand it, or want to learn it.

Analysis

The first exploratory item I did was visualizing a few images from the dataset. Not only did I want to ensure the data accuracy, but I wanted to see what resolution I was working with. It was hard to do any summary statistics or heatmaps, etc. due to the data being flattened images so no EDA would have been useful in this case.

However, the images have a 28 x 28 x 1 size, a very low resolution image of hands depicting each letter of the american alphabet. Since the images have been flattened into columns of greyscale values, there will be 784 columns, and all values except for the predictor variable were scaled between 0-1. The predictor variable is a single number that denotes a letter of the alphabet, excluding J and Z which require motion.

(27455, 784) (27455, 24)

(7172, 784) (7172, 24)

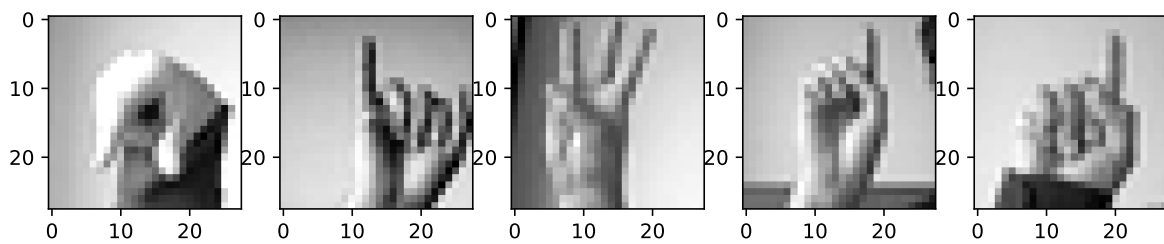
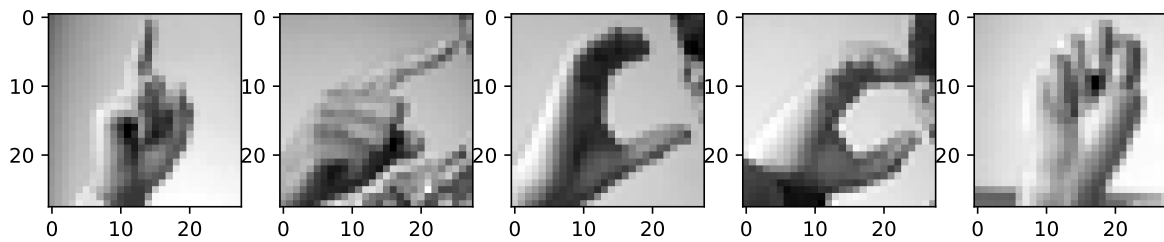


Figure 1: Random Images from Dataset