Modelling ASL Alphabet in Lower Dimensions

XueShan Bai, Vibhu Jawa, Praateek Mahajan, Xi Wang

Johns Hopkins University | Whiting School of Engineering | Baltimore, MD

Introduction

Figure 1 - Letter A in ASL Alphabet

The **objective** of our project is model the alphabets of the language to a lower dimension and see how the classification techniques perform in different latent space.

The results from this process can help perform better computer vision tasks at smaller resolutions, with very limited computing power. One could implement computer vision in an inexpensive board computer like Raspberry Pi, and some Text-to-Speech to enabling improved and automated translation applications.

Data Exploration and Preparation

- Each example in the dataset represents a label (0-25) as a one-to-one map for each alphabetic letter A-Z. However there are no cases for J/Z because they require gesture motions.
- The images in test set are distinct (clicked in different environment) from the images in train set. This help us get an estimate of how well does our model generalize.
- All the images in the dataset are in grayscale with and only 28 x 28 in dimension. We scale all our images to have zero mean and unit standard deviation.

Achievable Accuracies

The output of the algorithm is the classification accuracy achieved on 24 classes.

- A random classifier would score nearly 4%.
- A majority classifier would label everything E, thus giving 11%.

All our models perform better than these models, with the worst model giving accuracies in the range of 85%, after dimensionality reduction.

Methods and Models

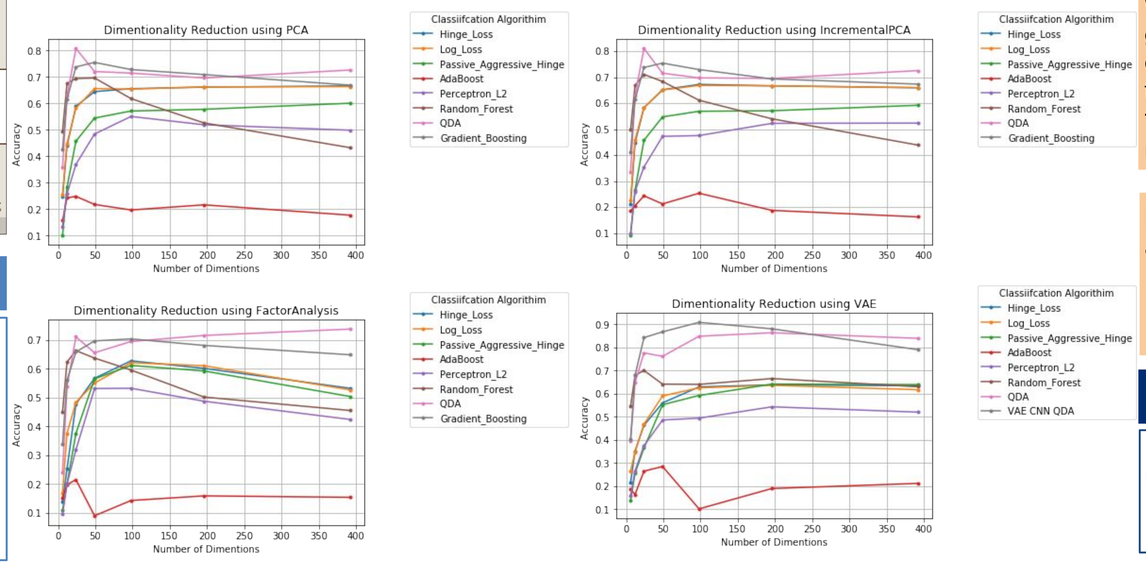
- We utilized multiple classifiers include **SGD Classifier using both Hinge Loss and Log Loss, Gradient Boosting, AdaBoost, Passive Aggressive, Perception, Random Forest and QDA.**
- We use several traditional techniques like **PCA, PPCA, Incremental PCA** for representing our 784 dimension image to a latent dimension of **392, 196, 98, 49, 24, 12, 6**.
- While these methods help us achieve a reduced dimensions statistically, we also explore newer methods such as **Variational AutoEncoders.**
- Due to high compute required by our tasks (224 experiments), we do limited grid search and hope that our results can be generalised for a combination of Classifier and Dimensionality Reduction Technique.

Empirical Results

Various Accuracy Results BEFORE Different Dimensionality Reduction Techniques

Classifier Methods	Train Accuracy
Hinge Loss	0.405
Log Loss	0.531
AdaBoost	0.120
Passive Aggressive - Hinge	0.532
Perception L2	0.477
QDA	0.860

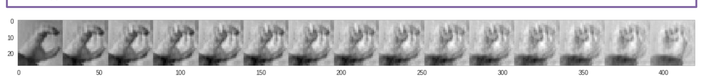
Various Accuracy Results AFTER Different Dimensionality Reduction Techniques



Variational AutoEncoders (VAE)

The top row represents the orgInal images and the bottom row is the image reconstructed by the decoder of vae using the latent representation.

Variational autoencoders (VAEs) are a deep learning technique for learning latent representations.
We present the images reconstructed from the latent space learnt by the auto encoder. It is clear from the reconstruction of images that we don't see much loss of information in the reduced dimensionality space We also showcase the transition b/w two labels using the hidden states, a capability which we gain because of the continuous nature of the latent space.
We can think of this transition as moving along the manifold of image features (think thickness, orientation, hand etc.)



Conclusion

Summary

We see that we can effectively reduce dimensionality and obtain good classification (test-90%) accuracy while reducing the dimensionality of data. The original data set was 786 using vae to further reduce it down to 98.

The overall 87.53% decrease in dimensionality makes viable realtime classification when running can on mobile devices.

Future Work

- Predictions about different ASL gestures can be made by various accurate models. Thus we can explore the possibility to include a camera to take live records of subject's gestures to enhance the his/her learning experience.

Acknowledgements

1. 553.636 Data Mining Lecture Notes, Professor Tamás Budavári
2. Gerard Aflague Collection, American Sign Language (ASL) Alphabet (ABC) Poster, 2015