Universal Classification Model Via Sparse Representation

Pavan Sudhir Nallam
Center for Data Science
New York University
New York, NY
psn240@nyu.edu

Abstract

In this project we want to replicate how the brain learns representations through vision. If we consider brain as single model, it looks at different classes of objects and trains the visual pathways to a classify different objects from different datasets. We like to extend the scope of the classification model proposed in (2) by extending the feature extraction to represent multiple datasources at the same time. We will train the model with a combined datasets consisting of faces and MNIST datasets.

1 Introduction

1.1 Motivation and Background

Replicating brain functionality has been a challenge for a long time. Human brain can grasp different formats of data like images, sounds and language etc., and process the data and use the data for predicting new ones.

In this project we concentrate on transfer learning, where we are utilizing just one format of data, images, and see if we can combine different objects and see model can distinguish them. As (6) suggests different kinds of data are in different feature space, so we need to make sure the feature space we pick can represent all those objects. For example, if we want to combine faces and MNIST data, the feature space should be sufficiently large to represent both the datasets.

1.2 Problem Definition

Each dataset has it's unique feature space,which can be used to represent it. Them more complex the data is the cardinality of the feature space increases. So we consider separate subspaces to represent each type of data, the feature space becomes huge and the problem becomes intractable. So we need to share the feature space, which uses minimum cardinality to represent multiple datasetrs. So (2) proposes sparse feature space to represent. As (2) suggests we would select sparse representation for the multiple datasets. We will concentrate more on feature extraction part of the problem.

As quoted in (2) 'the theory of compressed sensing implies that the precise choice of feature space is no longer critical: even random features contain enough information to recover the sparse representation and hence correctly classify any test image. What is critical is that the dimension of the feature space is sufficiently large, and that the sparse representation is correctly computed', we can select l^1-Minimization for getting the sparse representation of the data.

2 Related Work

There has been many studies out there, for single model to represent different kinds of data. Recently google brian team released a paper (7), where they have taken different formats of data and used convolutional layers, an attention mechanism, and sparsely-gated layers to represent them. (8) has used multi-task approach for Natural language processing models. Multi-modal learning has been shown to improve learned representations in the unsupervised (9).

Learning feature space from one datasource and reapplying that model to another datasource is called transfer learning. In this approach we need not relearn the domain space of the new datasource. As Survey on Transfer learning (6), gives a overall picture about transfer learning and essential details about the implementations. Transfer learning helps in label minimization.

3 Data

In this work we have picked two datasets, each dataset are images with varying complexity in the features. All the images are converted to grayscale.

3.1 Extended Yale B Database

The Extended Yale B database consists of 2,414 frontal-face images of 38 individuals and around 64 near frontal images under different illuminations per individual (3). The cropped and normalized 192x168 face images were captured under various laboratory-controlled lighting conditions (5).

Figure 1: Sample images from Extended YaleB dataset

3.2 MNIST Dataset

We have taken MNIST datset from the class assignment as the toy dataset. MNIST has hand written digits from 0 to 9 as images. There are 1000 images in training with 100 images for each class. Since the feature space to represent it is small, we have experimented by combining with other datasets.

Figure 2: Sample images from MNIST dataset

4 Approach

Basic approach proposed by (2) is representing the test sample as linear combination of the training data. In this paper we extend idea by combining the training and testing data of two class of images together and represent the test image as sparse representation of the training images.

Given sufficient training samples of the i*th* object class, $A_i = [v_{i,1}, v_{i,2}, v_{i,3} .... v_{i,n_i}]$, any new (test) sample $y \epsilon R^{mxn}$from the same class will approximately lie in the linear span of the training samples associated with object i:

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \alpha_{i,3} v_{i,3} + ... + \alpha_{i,n_i} v_{i,n_i}$$

$$y = Ax$$

There are two ways of solving the problem, one using least square to solve $||Ax - Y||_2$ given by,

$$\hat{x} = argmin||x||_2$$

subject to Ax = y

The solution we get from this is not very informative, since the weights of almost all the training data are non-zero, there by giving very dense solution. Second of solving it is by,

$$\hat{x} = argmin||x||_1$$

subject to Ax = y
This equation can be solved using linear programming. The solution we get from this sparse enough, there by giving an approach better approach then l_2 norm. Fig. 3 gives a geometric interpretation of why minimizing the l_1 norm correctly recovers sufficiently sparse solutions.

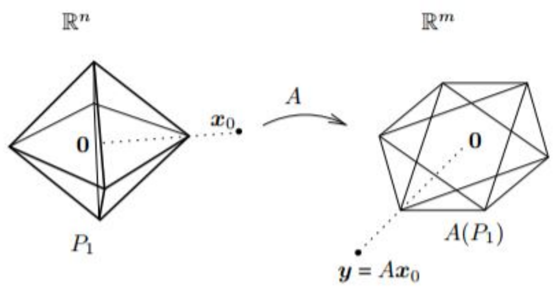

Figure 3: Geometry of sparse representation via l_1-minimization, figure from []

4.1 Classification model based on sparse representation[1]

We have followed the same implementation algorithm stated in the (2), instead of one datasource, we have combined two or three datasources. Pseudo code is shown below in the figure 4.

1: **Input:** a matrix of training samples $A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{m \times n}$ for k classes, a test sample $\boldsymbol{y} \in \mathbb{R}^m$, (and an optional error tolerance $\varepsilon > 0$.)
2: Normalize the columns of A to have unit ℓ^2-norm.
3: Solve the ℓ^1-minimization problem:
$$\hat{\boldsymbol{x}}_1 = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad A\boldsymbol{x} = \boldsymbol{y}.$$
(Or alternatively, solve
$$\hat{\boldsymbol{x}}_1 = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \|A\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \varepsilon.)$$
4: Compute the residuals $r_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\,\delta_i(\hat{\boldsymbol{x}}_1)\|_2$ for $i = 1, \ldots, k$.
5: **Output:** $\text{identity}(\boldsymbol{y}) = \arg\min_i r_i(\boldsymbol{y})$.

Figure 4: Sparse Representation based classification, algorithm from []

5 Results

First we have tried to replicate the results on the two datasets. And then we can combined the two datasets and tried to classification with the combined data. The results are presented according to the data sets below.

[1] Code: `https://github.com/sudhirNallam/sparse_representation`

5.1 MNIST Dataset

For MNIST dataset, figure 5 shows the test case used for testing the classification model. And figure 6 shows the weight distribution over the training data.

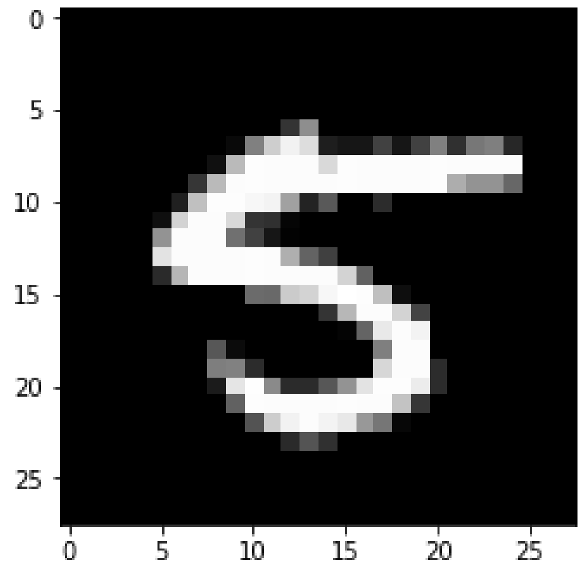

Figure 5: MNIST test case

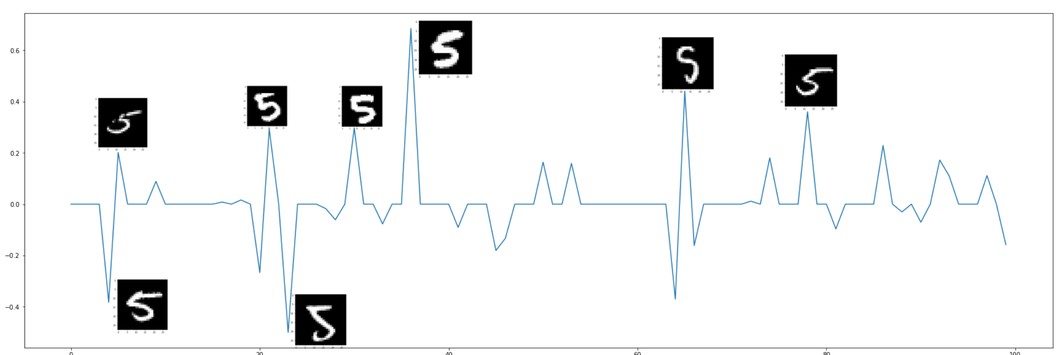

Figure 6: Weight distribution for MNIST dataset with MNIST test case

5.2 Faces Dataset

For Faces dataset, figure 7 shows the test case used for testing the classification model. And figure **??** shows the weight distribution over the training data.

5.3 Combined Dataset

For the combined datasets, we have taken only 50 images for each category for the MNIST dataset. When we trained the model with MNIST test case the weights for all the faces training data were zeroes. Similarily when we trained the model with Faces test case the weights for all the MNIST traning data was zeroes. Figure 9 shows the one of the weight distribution for the MNIST training data. And figure 10 shows the one of the weight distributions of the faces training data.

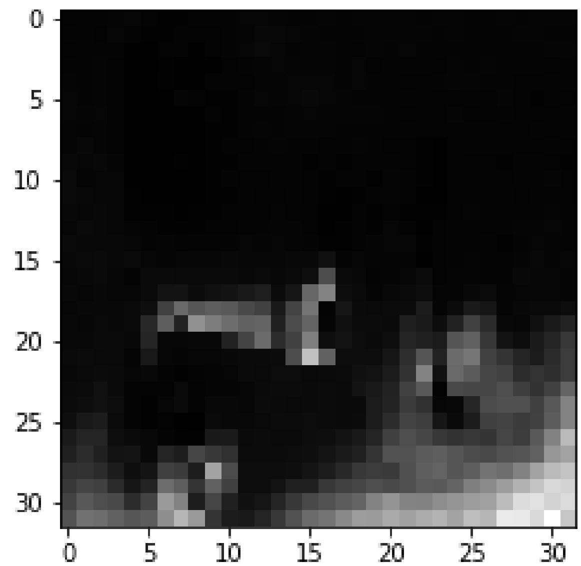

Figure 7: Faces test case

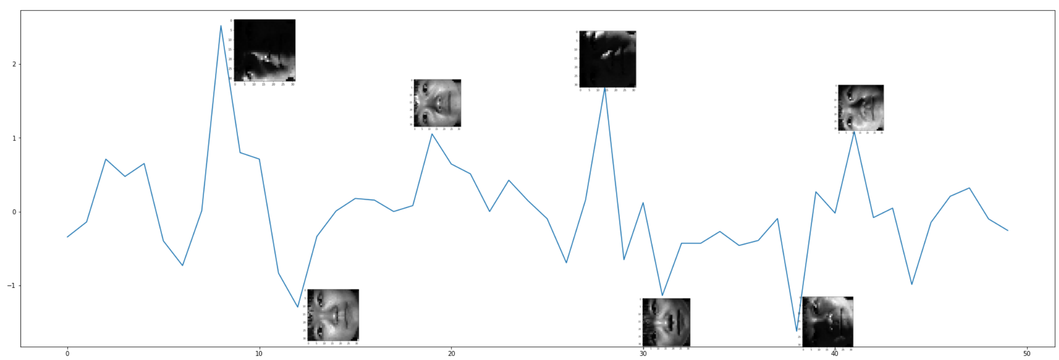

Figure 8: Weight distribution for Faces dataset with Faces test case

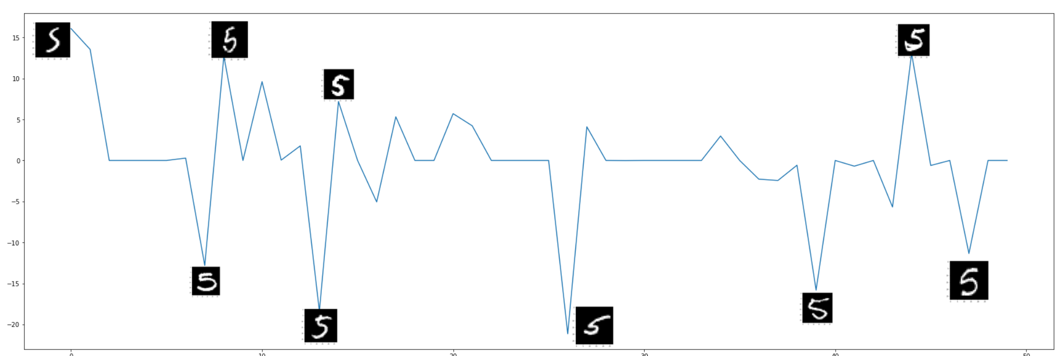

Figure 9: Weight distribution for Mixed dataset with MNIST test case

6 Conclusion

Universally learning a single model that represents multiple datasources has been goal for a long time. One of the limitations of that is feature space, which grows as we add datasources. In this

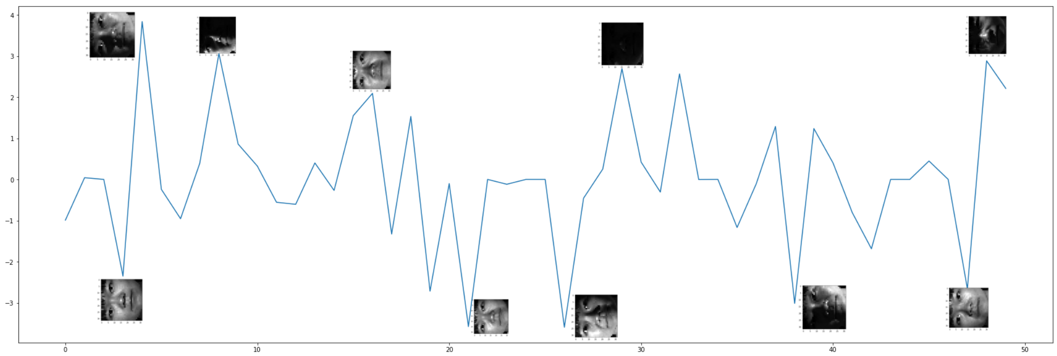

Figure 10: Weight distribution for Mixed dataset with Faces test case

project we have shown that by picking sparse representation of the feature space, can handle multiple datasources gracefully. Further we can extend the training process, by passing the feature space to a deep learning network.

References

[1] John Wright, Arvind Ganesh, Allen Y. Yang, Zihan Zhou and Yi Ma. *Sparsity and Robustness in Face Recognition*

[2] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma *Robust face recognition via sparse representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2):210 – 227, 2009.

[3] A. Georghiades, P. Belhumeur, and D. Kriegman. *From few to many: Illumination cone models for face recognition under variable lighting and pose*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 643–660, 2001

[4] K. Lee, J. Ho, and D. Kriegman. *Acquiring linear subspaces for face recognition under variable lighting*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 684–698, 2005

[5] Aditya Khosla and Nityananda Jayadevaprakash and Bangpeng Yao and Li Fei-Fei. *Novel Dataset for Fine-Grained Image Categorization*. First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition, 2011

[6] Pan, Sinno Jialin and Yang, Qiang. *A Survey on Transfer Learning*. IEEE Trans. on Knowl. and Data Eng., 2010

[7] Lukasz Kaiser, Aidan N. Gomez, Noam Shazeer, Ashish Vaswani, Niki Parmar, Llion Jones and Jakob Uszkoreit. *One Model To Learn Them All*. CoRR, 2017

[8] Ronan Collobert and Jason Weston. *A unified architecture for natural language processing: deep neural networks with multitask learning*. In Proceedings of the 25th International Conference on Machine learning, pages 160–167, 2008.

[9] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y. Ng. *Multimodal deep learning*. In Proceedings of ICML'11, pages 689–696, 2011.