# CSE 847 Project Proposal

Discovering the underlying topological space of extracted features

Hamid Karimi
karimiha@msu.edu

Harrison LeFrois
lefroish@math.msu.edu

**ACM Reference format:**
Hamid Karimi and Harrison LeFrois. 2017. CSE 847 Project Proposal. In *Proceedings of The Best Conference, East Lansing, MI, 04/28/2017,* 2 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

## 1 PROBLEM DESCRIPTION

For our project, we have decided to examine what is happening in the hidden layers of neural networks. Apart from the simplest examples, the maps that a neural network is learning become complicated very quickly. We are particularly interested to see if it is possible to find an underlying topological space (potentially a manifold) that parametrizes our feature space for a layer in our network. In other words, what is the space that spans our selected features? Is it possible to relate the representation of our data at each layer to something geometric? Data analysis generally does not leverage the geometry and topology of the data[1], which is what we are aiming towards. It may not be possible to find an underlying manifold for our data, but parameterizing the feature space by a topological space can help us better understand the particular data set and what is happening in our network.

## 2 SURVEY AND RELATED WORK

Dimensionality reduction is a well-established and often used strategy to analyze high dimensional data and enables the user to visualize data that would otherwise be inaccessible. This is done by extracting features that best represent the data and mapping the data to a lower dimensional space based on these features, hopefully without too much loss of information.

Some common methods of dimensionality reduction and manifold learning are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Multidimensional Scaling (MDS), and ISOMAP. PCA, LDA, and MDS are linear approaches that are not very useful for non-linear structures. While ISOMAP is a non-linear approach, it fails to retrieve the underlying embedding in complicated structures. Nevertheless, deep neural networks have shown to be quite successful in extracting the underlying manifold of data feature space.

One of the common networks for manifold learning is autoencoders. Autoencoders essentially seek to learn the identity map, i.e. they have the same number of input nodes as output nodes. However, by varying the number of nodes in the hidden layers (say less than the input layer) or imposing sparsity, one can extract a number of features which is less than the dimension of the data's ambient space. If the data was sampled from a manifold embedded in the ambient space, the hope is that the extracted features can tell us about the manifold and/or potentially allow us to reconstruct the manifold. This is the general goal of manifold learning. But what if the data is not sampled from a manifold? Then maybe it is possible to examine the extracted features and deduce an underlying space for the features. We intend to utilize sparse autoencoders during our investigation. Sparse coding has seen great success in feature learning [2], and our hope is that utilizing the geometry/topology of the feature space may yield new insights.

This project is inspired by the work of Dr. José Perea in [3]. A motivating example in his work involved finding the underlying topological space of $7 \times 7$ pixel grey-scale images displaying a fixed-width line. Let $X \subset \mathbb{X}$ denote a sample of these images from the set $\mathbb{X}$ of all such images. Each image could be represented as a $7 \times 7$ matrix whose entries, consisting of -1 for black and 1 for white, sum to 0. By concatenating the columns of each matrix, we could view this data set $X$ as a subset of $\mathbb{R}^{49}$. He noticed that despite being embedded in a 49 dimensional space, each image was governed by the distance of the line from the center and the angle of the line compared to the horizontal. Locally, $\mathbb{X}$ appears to be 2 dimensional. Using a combination of MDS and a strategy to discover the orientability of $\mathbb{X}$, Perea was able to conclude that $X$ was actually sampled from the real projective plane $\mathbb{R}\mathbf{P}^2$. Each image in $X$ was then able to be assigned a pair of coordinates in $\mathbb{R}\mathbf{P}^2$ by using an extension of PCA called Principal Projective Components. In this particular instance, the data was sampled from a manifold and the important features of each image, namely the angle and distance, provided some of the information needed to find the underlying topological space. The goal of this work, possibly in conjunction with algebro-topological tools, is to utilize machine learning to make similar discoveries.

## 3 PLAN AND MILESTONES

With an eye towards Dr. Perea's work, we would like to see if autoencoders can provide a way towards parameterizing the feature space of other data sets. The data set he used provides an opportunity to test and see if an autoencoder can detect features that would hint at $\mathbb{R}\mathbf{P}^2$ being the underlying topological space. We will utilize different sparse autoencoder architectures to extract features from our data sets to see if each autoencoder focuses on different features (and hence a potentially different space spanning the features). Using data from [3], MNIST, and potentially natural images, we will apply our algorithms to each data set and attempt to find an underlying topological space that would parameterize the feature space. Our projected timeline and milestones are in the following chart.

*The Best Conference, East Lansing, MI*
2017. 978-x-xxxx-xxxx-x/YY/MM…$15.00
DOI: 10.1145/nnnnnnn.nnnnnnn

| | |
|---|---|
| **Week 1-2 (02/17-03/03)** | Literature survey and complete sparse autoencoder tutorial [4] |
| **Week 3-4 (03/03-03/17)** | Construct autoencoders and test on data from [3]. Prepare Intermediate Report |
| **Week 5-6 (03/17-03/31)** | Run autoencoders on data from MNIST and natural images. Analyze feature space. |
| **Week 7-11 (03/31-04/28)** | Continue feature space analysis. Prepare final report. |

## REFERENCES

[1] Woong Bae, Jae Jun Yoo, and Jong Chul Ye. Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. *CoRR*, abs/1611.06345, 2016. URL http://arxiv.org/abs/1611.06345.

[2] Yunlong He, Koray Kavukcuoglu, Yun Wang, Arthur Szlam, and Yanjun Qi. Unsupervised feature learning by deep sparse coding. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 902–910. SIAM, 2014.

[3] J. A. Perea. Multi-Scale Projective Coordinates via Persistent Cohomology of Sparse Filtrations. *ArXiv e-prints*, December 2016. URL http://adsabs.harvard.edu/abs/2016arXiv161202861P.

[4] Andrew Ng, Jiquan Ngiam, Chuan Yu Foo, Yifan Mai, and Caroline Suen. UFLDL Tutorial, 2013. URL http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial.