Multilayer NMF: An Experiment

Prince Zizhuang Wang
University of California, Santa Barbara
College of Creative Studies

December 16, 2017

Abstract

The key idea of deep learning is to learn a set of representational features that are embedded in lower dimension space than the original input space. However, little is known about the changes of internal states of hidden layers as training goes on, nor do we understand the actual interpretation of the learned features. Hence, in this experiment, we propose a new deep learning model which consists layers of nonnegative matrix factorization to learn sparse and meaningful representations. Nonnegative matrix factorization(NMF) is a computational technique for dimensional reduction of given data to uncover the latent factors embedded in higher dimensional space. The two key features of NMF are: (1) the learned representations are interpretable, because with the nonnegative constrains posed on the coefficients matrix, features are combined additively to reconstruct the original data, (2) the learned feature map can be very sparse. In the experiment, we applied a three layer NMF network to learn the latent features embedding in the sample images, and we show that the reconstructed image generated from these latent features matches to the original one, which consolidates our belief in the expressiveness of the features learned by simple matrix decompostition.

1 Introduction

Nonnegative matrix factorization is a computational technique of dimensional reduction of a given data to uncover the latent factors embedded in higher dimensions. Unlike traditional matrix decomposition methods such as SVD and full rank decomposition, the non-negativity constraint imposed by NMF is useful for learning part-based representations. Secondly, since (1)the data matrices people are dealing with in many real world applications such as image and face recognition are usually nonnegative, (2) intuitively an object is made by many parts of it whose information about the original object is not negative, (3)and that physiological principles show that human learn objects from part to part, NMF network therefore gives meaningful interpretation of inputs from the viewpoint of information theory and is applicable to many real world problems.

Definition 1.1 *Nonnegative Matrix Factorization is formulated as the following: for any given matrix $X \geq 0$, we want to find a pair of matrix F, G, such that the Frobenius norm $||X - FG^T||$ is minimized. And we do this by solving the following optimization problem*

$$min||X - FG^T||^2$$

$$subject\ to, F \geq 0, G \geq 0$$

In Semi NMF, we clear the nonnegative constraint on matrix F so that (1) we can decompose any given matrix, rather than just nonnegative ones, (2) the learnable features can be negative and therefore the volume of the convex cone spanned by features can be larger, being able to represent more data points.

2 Experiments

We first feed the sample images into a one layer NMF. The result matrices can be viewed as features and their coefficients, and the original input is simply the multiplication of these two matrices. By viewing the feature matrix as a combination of column vectors, and the coefficient matrix as a combination of row vectors, we see that the original image can be interpreted as the summation of the outer product between each feature vector and each coefficient vector, which is equivalent to say that it is a linear combination of features. In order words, the feature vectors learned by NMF span the vector space that contains the original input image.

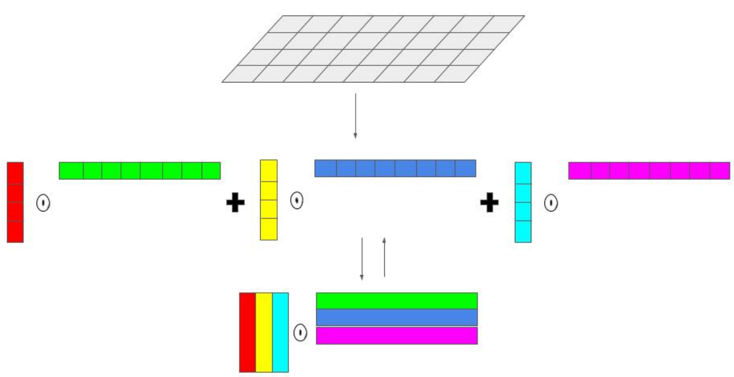

Figure 1: NMF layer. It can be viewed as the summation of the outer product between each feature vectors and each coefficient vectors. The matrix consisting of column vectors is the feature map, and the one made by row vectors is the coefficient matrix.

According to our observation, when we feed the network with different sample images, the coefficient matrix also changes enormously, while the features do not change too much. Such phenomenon resembles the behavior of network of neurons in human brain. In 'Dynamic Routing between Capsules', Hinton argues for the equivariance instead of invariance property of the current state-of-art convolutional neural network and its variants. A network is invariant if its neural activities do not change a lot when it is given the same input object but only from different viewpoints, meaning that the knowledge of this object is encoded in the pattern of neurons' firing rate, and it is equivariant if the neurons behave differently but the synapses connecting them stay static, meaning that the knowledge is stored in the synapses, namely, the weights. And if the goal of artificial neural network is to mimic the working mechanism of neurons, then it should be equivariant because the neuron activities in human brain change a lot even when it is recognizing the same objects but only from different viewpoints [1].

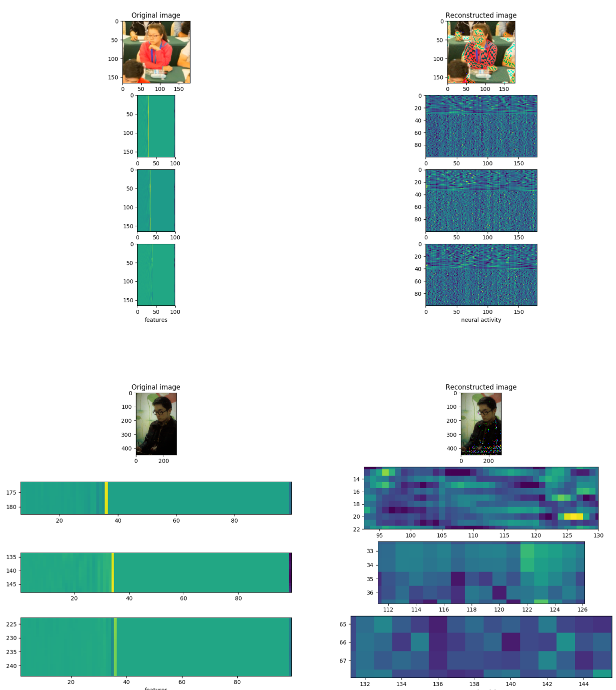

Figure 2: Features and representations learned by one layer NMF network. Images reconstructed from the features match to the original ones precisely. We also observe that the features of images of different objects have the same pattern, while the neural activities change enormously when it is given different objects.

To further test whether NMF network is equivariant, we feed it with completely different images, each of which consists of pixels randomly generated.

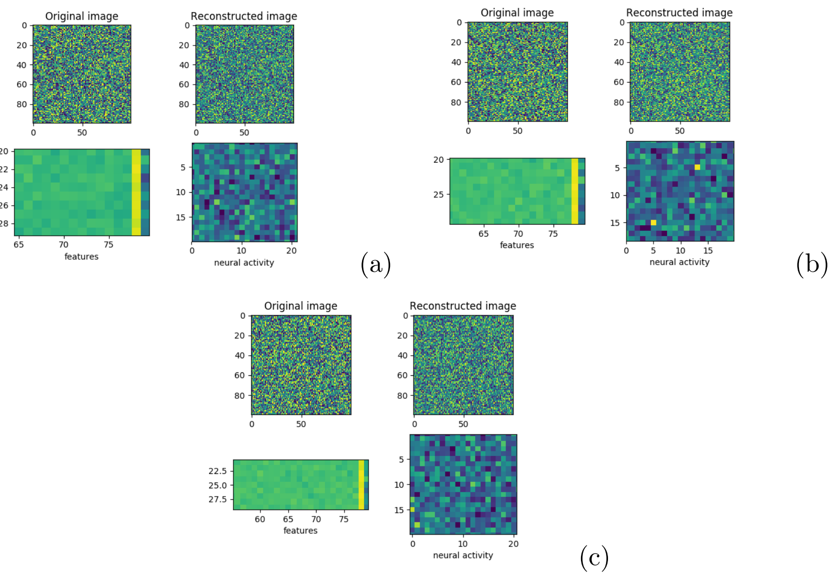

Figure 3: the original images in (a),(b),(c) are all generated with random pixels. It is obvious by comparison that even with random images, the patterns of features are still very similar to each other, while the neural activities are completely different.

As in Figure 3, if we compare the features and representations associated with each random images, then it is obvious that the patterns of features of different inputs are strongly related to each other, but their neurons' activities are just as random as the original images, which further consolidates our belief that NMF is equivariant and hence it stores knowledge in the weights as what neuron does.

3 Multilayer NMF network

A multilayer NMF network is a sequential model that consists of many layers of NMF. At each layer, the learned representation is fed into the next layer, so that each layer learns a latent, low dimensional representations of the original ones, along with higher-level abstract features that are more expressive.

For simplicity of our presentation, we applied a three layer NMF network to our sample image. The network learns features embedding in a lower dimensional space than that learned by one layer network. However, we noticed that reconstruction error, the difference between the original image and the reconstructed one, is larger than that of the shallower network. This is due to the inevitable information loss of any feed-forward network structure. To resolve the problem of information loss of NMF networks, further researches on more exact approximation of nonnegative matrix factorization problem are needed.

3.1 Project Site

https://github.com/kingofspace0wzz/multilayer_nmf

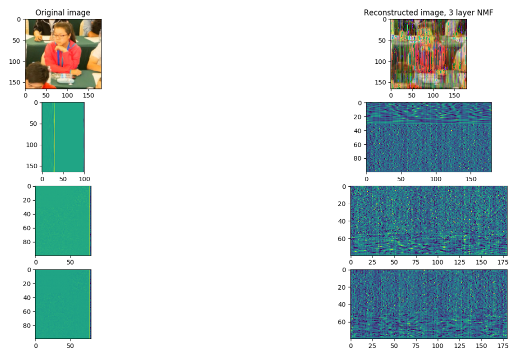

Figure 4: Results from a three layer NMF network

References

[1] Sara Sabour. & Nicholas Frosst. & Geoffrey E Hinton. (2017) Dynamic Routing Between Capsules. *https://arxiv.org/abs/1710.09829*.