Training Autoencoders in Sparse Domain

Biswarup Bhattacharya
University of Southern California
Los Angeles, CA 90089. USA.
Email: bbhattac@usc.edu

Arna Ghosh [*]
McGill University
Montreal, QC H3A 0G4. Canada.
Email: arna.ghosh@mail.mcgill.ca

Somnath Basu Roy Chowdhury [*]
Indian Institute of Technology
Kharagpur, WB 721302. India.
Email: brcsomnath@ee.iitkgp.ernet.in

Abstract

Autoencoders (AE) are essential in learning representation of large data (like images) for dimensionality reduction. Images are converted to sparse domain using transforms like Fast Fourier Transform (FFT) or Discrete Cosine Transform (DCT) where information that requires encoding is minimal. By optimally selecting the feature-rich frequencies, we are able to learn the latent vectors more robustly. We successfully show enhanced performance of autoencoders in sparse domain for images.

Introduction

Autoencoders play an important role in machine learning. They are simple learning models which aim to transform inputs into outputs with the least possible amount of distortion. Autoencoders were first introduced by (Rumelhart et al. 1988) to address the problem of "backpropagation without a teacher", by using the input data as the teacher. The biological essence of AEs lies in the fact that AEs are a proxy for addressing the mystery of how synaptic changes induced by local biochemical events can be coordinated in a self-organized manner to produce global learning and intelligent behavior. More recently, AEs have gained prominence due to the applicability of deep architectures in their design which has led to state-of-the-art results.

In neural net language, a variational autoencoder consists of an encoder, a decoder, and a loss function. A typical autoencoder can usually encode and decode data very well with low reconstruction error, but the latent code does not learn the probability distribution of the data. Therefore, if we are interested in generating more data, a typical autoencoder generally do not seem to work as well. Another version of autoencoders, called "variational autoencoder (VAE)", is used to solve this problem since they explicitly define a probability distribution on the latent code. Even though the actual neural network architecture is very similar to a regular autoencoder, the difference is that the hidden code comes from a probability distribution that is learned during the training. VAEs were defined by (Kingma and Welling 2013) and (Rezende, Mohamed, and Wierstra 2014).

[*]Equal contribution

Autoencoders are essential in learning representation of large data (like images) for dimensionality reduction. (Hinton and Salakhutdinov 2006) described an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis (PCA). In this paper, we convert the input images to a sparse domain using transforms like Fast Fourier Transform (FFT) or Discrete Cosine Transform (DCT). Fourier analysis converts a signal from its original domain (usually time or space) to a representation in the frequency domain (and vice versa). A Fast Fourier Transform (FFT) rapidly computes such transformations by factorizing the Discrete Fourier Transform (DFT) matrix into a product of sparse factors, thereby, reducing the complexity of computing the DFT.

In our case, the images (originally in spatial domain) are converted to frequency domain using the transform techniques. Dominant spatial frequencies (which define main structure or edges) are located in the center of the transformed image while frequencies away from the center contain finer details of the image. Using autoencoder if we are able to capture the central spatial frequencies it would aid in reducing loss to a significant amount, without having to consider the entire image. This reduces data samples the autoencoder needs to process to learn an embedding in reduced space. We are able to show enhanced performance of autoencoders in sparse domain for images.

Motivation

Autoencoders use neural networks to encode and decode data distributions. It is often observed that autoencoders end up memorizing data points and are not able to learn an efficient latent vector of the data. Autoencoder architectures with high number of parameters, map individual data points to discrete points completely ignoring the correlation among them. This phenomenon makes the performance of an autoencoder heavily dependent on the power (number of layers) of the decoder. Researchers are trying to figure out an autoencoder with a robust decoder. We use our autoencoder module to map the FFT of closely linked images - alike images from same distribution to map to similar distribution. This is inspired by the fact that motion cues or temporal cues act as weak supervisory signals for object detection (Agrawal, Carreira, and Malik 2015), (Goroshin et al. 2015).

Architecture

We use a simple two-layered autoencoder network, although the proposed model is applicable to other stacked autoencoder models. The network reduces the dimension of MNIST dataset from 784 to 20. Table 1 depicts the entire network architecture. Layers 1 & 2 in Table 1 are the encoder module and Layers 3 & 4 constitute the decoder module.

Layer	Type	Maps and Neurons
1	Linear	784 × 400
2	Linear	400 × 20
3	Linear	20 × 400
4	Linear	400 × 784

Table 1: Autoencoder Network architecture

In our experiments we also deploy a single layered decoder of dimension 20 × 784, for robustness comparison.

Experiments

Setup

We use the MNIST handwritten digits dataset for experimentation. We train the network on the MNIST images. The training set has 60000 images and the test set has 10000 images. The network is trained using Adam Optimizer with a learning rate of 0.001. The batch size is kept at 256 images and the network is trained for 20 epochs.

Results & Illustrations

The test-set (magnitude reconstruction) loss of AE in sparse domain and vanilla AE is shown in Figure 1. We observe that the AE in Fourier domain outperformed the vanilla AE model in terms of binary cross-entropy loss. The proposed network converged exceedingly well, reaching near optimal value after the first epoch only. In Figure 2, we have also shown the test-set loss when 1 layer of decoder was removed.

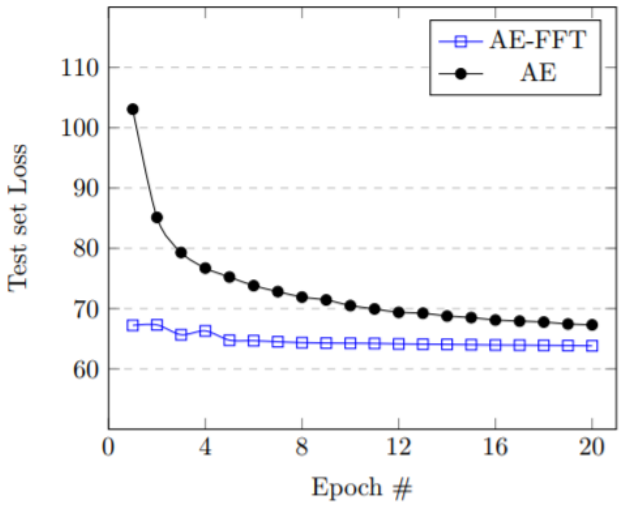

Figure 1: Test-set Loss for 2-layered AE-FFT and AE

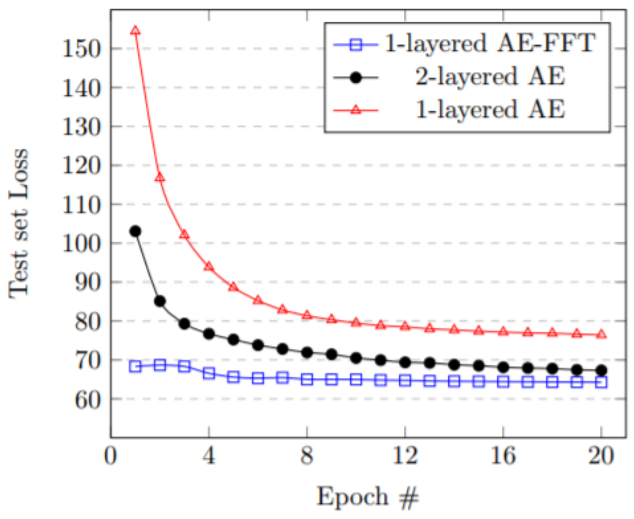

Figure 2: Test-set Loss for AE-FFT, 1-layered vanilla AE & 2-layered vanilla AE

Figure 2 shows the test-set (magnitude reconstruction) loss for AE & AE-FFT having a single decoder layer (20×784). The two losses are also compared with the vanilla AE model have 2-layered decoder. From the plot we can observe that a 1-layered AE-FFT outperforms the vanilla 2-layered model. Also it is seen that AE-FFT is robust to decoder layer reduction & increase in cross-entropy loss is much less compared to vanilla AE.

Conclusion

We show that by optimally selecting feature-rich frequencies, we are able to enhance performance of prevalent Autoencoder architectures. This reduces the dependency of autoencoder architectures on huge datasets. The test-set loss of AEs in sparse domain outperforms the vanilla AE model thereby showing enhanced performance of autoencoders in sparse domain for images.

References

Agrawal, P.; Carreira, J.; and Malik, J. 2015. Learning to see by moving. In *Proceedings of the IEEE International Conference on Computer Vision*, 37–45.

Goroshin, R.; Bruna, J.; Tompson, J.; Eigen, D.; and LeCun, Y. 2015. Unsupervised feature learning from temporal data. *arXiv preprint arXiv:1504.02518*.

Hinton, G. E., and Salakhutdinov, R. R. 2006. Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507.

Kingma, D. P., and Welling, M. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Rezende, D. J.; Mohamed, S.; and Wierstra, D. 2014. Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*.

Rumelhart, D. E.; Hinton, G. E.; Williams, R. J.; et al. 1988. Learning representations by back-propagating errors. *Cognitive modeling* 5(3):1.