Auto-Encoder

Subham Roy
Department of Computer Science and Engineering
University at Buffalo, Buffalo, NY 14260

1 Overview

Auto-Encoder is an unsupervised artificial neural network that learns how to efficiently compress and encode data then learns how to reconstruct the data back from the reduced encoded representation to a representation that is as close to the original input as possible. Autoencoder, by design, reduces data dimensions by learning how to ignore the noise in the data.

2 About Auto-Encoder

Autoencoders consists of 4 main parts:
1- Encoder: In which the model learns how to reduce the input dimensions and compress the input data into an encoded representation.
2- Bottleneck: which is the layer that contains the compressed representation of the input data. This is the lowest possible dimensions of the input data.
3- Decoder: In which the model learns how to reconstruct the data from the encoded representation to be as close to the original input as possible.
4- Reconstruction Loss: This is the method that measures measure how well the decoder is performing and how close the output is to the original input.
The training then involves using backpropagation in order to minimize the network's reconstruction loss.

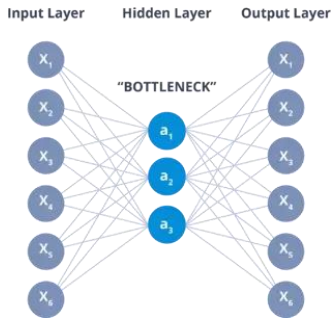


Fig: Layer structure of Auto-Encoder

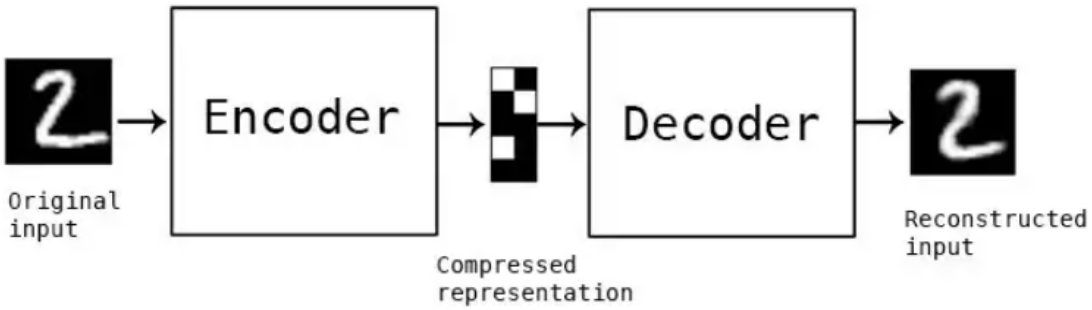


Fig: Autoencoder for MNIST

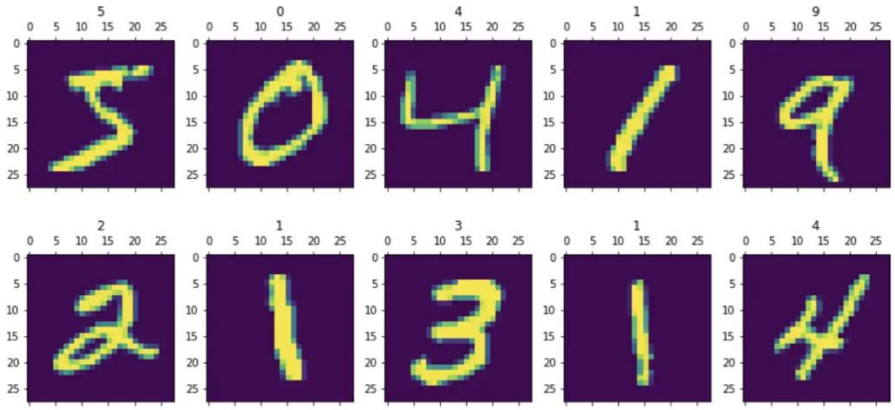
The network architecture for autoencoders can vary between a simple FeedForward network, LSTM network or Convolutional Neural Network depending on the use case.

3 Experiment

For our analysis, we are using MNIST handwriting digit dataset. We have approximately 60,000 images in the training set and around 10,000 images in the test set. All the images are of 28x28 pixels. Since Autoencoder is an unsupervised learning algorithm, there are no target columns. Therefore, all the columns in this dataset will be treated as predictors which are also known as dependent variables. This

```
X_train shape: (60000, 28, 28)
Y_train shape: (60000,)
X_test shape: (10000, 28, 28)
Y_test shape: (10000,)
```

dataset does not have any null values. The data is a ndarray. Sample output for the dataset depicting the images as numbers.



4 Analysis

In the below figure, we are showing the first 10 digits of the dataset. Now we will be adding some noise to the dataset. After that we will be removing this added noise using denoising autoencoder and then we will compare all these datasets.

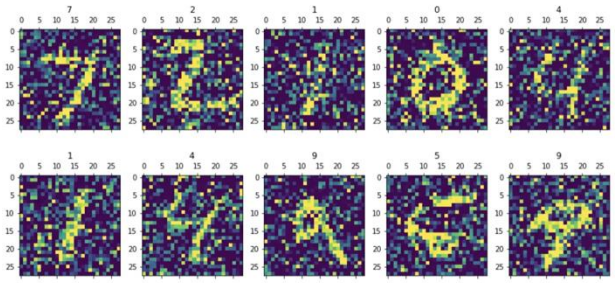


Fig: Dataset with noise

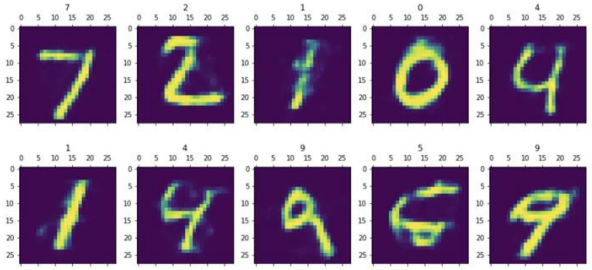


Fig: Denoised the dataset using Auto-Encoder

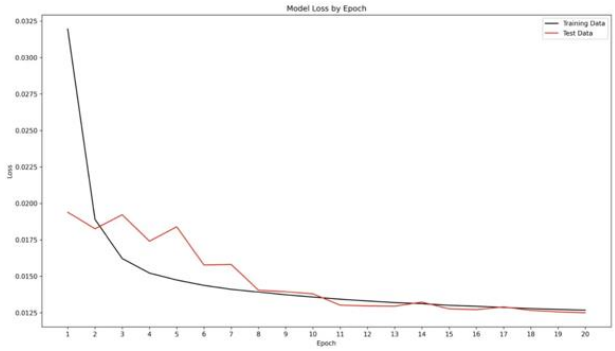


Fig: Denoising autoencoder model loss by epoch

References

[1] Gareth James [at], Daniela Witten [aut], Trevor Hastie [aut, cre], Rob Tibshirani [aut], Balasubramanian Narasimhan [ctb], *Introduction to Statistical Learning, Second Edition*.

[2] Tom M. Mitchell, *Machine Learning: A multistrategy approach, 1997 Edition*.

[3] URL - https://github.com/udacity/deep-learning-v2-pytorch/tree/master/autoencoder; https://towardsdatascience.com/auto-encoder-what-is-it-and-what-is-it-used-for-part-1-3e5c6f017726