# Extracting and Composing Robust Features with Denoising Autoencoders

Sanika Shah, Arizona State University
ssshah33@asu.edu, 1215159530

## I. Introduction

Autoencoders are a specific type of feedforward neural networks where the input is the same as the output. They compress the input into a lower-dimensional code and then reconstruct the output from this representation. We will implement Denoising autoencoder and Stacked autoencoder. Denoising autoencoders take input that has noise to it and recover the original noise-free data. A stacked autoencoder is a neural network consisting of multiple layers of sparse autoencoders in which the outputs of each layer is wired to the inputs of the successive layer. We will be using 3 layers excluding the input layer. We will use fashion-MNIST dataset to train and test our models. We aim to implement the neural networks from scratch using numpy, without the use of any other existing libraries. This project has two parts:
(i) De-noising autoencoder
(ii) Stacked autoencoder

### A. De-noising autoencoder

we will train an autoencoder with one hidden layer to reconstruct de-noised images from noisy versions. We will also keep varying the noise levels in our input to evaluate the de-noising capabilities of our network.

### B. Stacked autoencoder

We will train an autoencoder layer-by-layer in an unsupervised manner. We will stack 3 layers excluding the input layer to create a stacked autoencoder, and train it to act as an unsupervised feature extractor. We will also fine-tune the stacked autoencoder with labeled samples. by considering (i) 1-labeled sample per class, (ii) 5-labeled samples per class. We will evaluate the classification accuracies on the test dataset using these models.

## II. Related Work

### A. Denoising autoencoder

Denoising autoencoders take a partially corrupted input while training to recover the original undistorted input. This technique has been introduced with a specific approach to good representation in [1] paper. A good representation is one that can be obtained robustly from a corrupted input and that will be useful for recovering the corresponding clean input.

### B. Stacked autoencoder

## III. Method

## IV. Experiments

## V. Conclusions

## VI. Division of Work

## VII. Self-Peer Evaluation Table

## References

[1] - P Vincent, H. Larochelle Y. Bengio and P.A. Manzagol, Extracting and Composing Robust Features with Denoising Autoencoders, Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML08), pages 1096 - 1103, ACM, 2008.