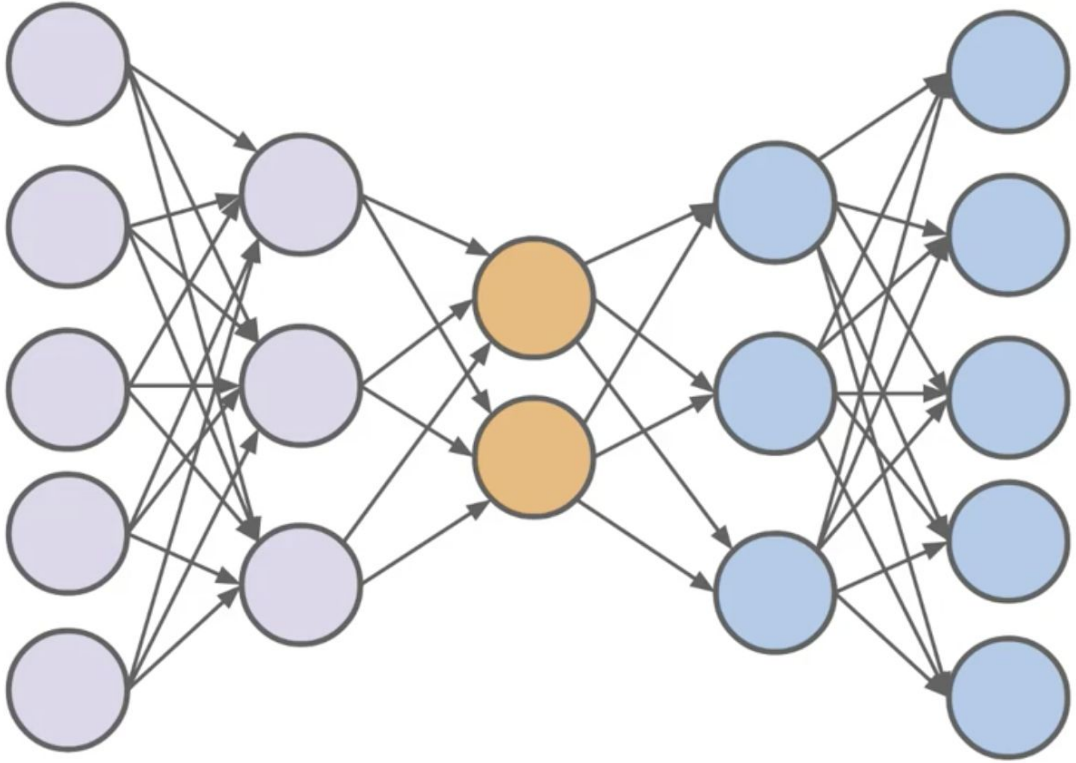
Autoencoders

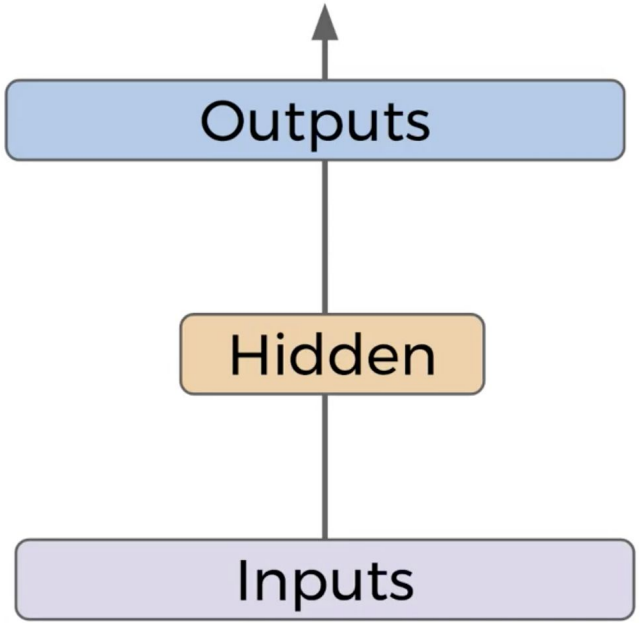
What it is?

- Simple neural network, essentially a MLP with a variation on the topology
- It is designed to reproduce it’s input in the output layer - Equal number of input and output neurons

General Concepts

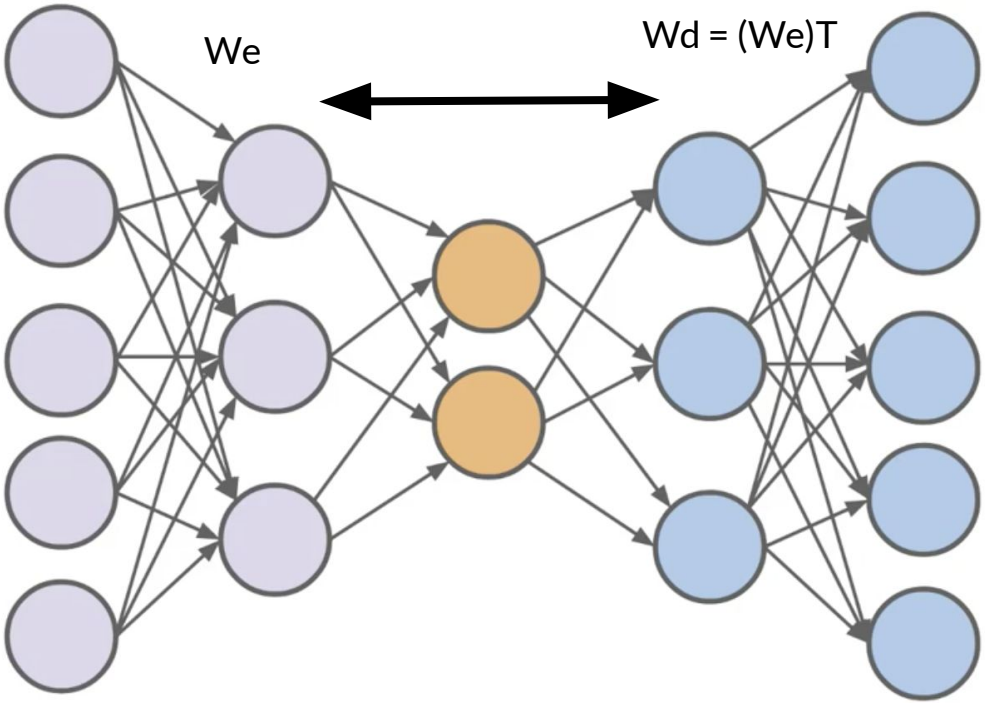


Simplified View



- Just an input layer attempting to re-create itself at the output layer
- **Encoder**: takes information and reduces it to a smaller representation
- **Decoder**: takes smaller representation and attempts to re-create the input
- This means learning and internal, compressed representation. Learning the important stuff
- Mimic PCA

Tied Weights



*Not biases, just weights

Linear Autoencoder

- Simple data compression
- 3 - 2 - 3
- Just simple linear transform, without activation
- Interpret results

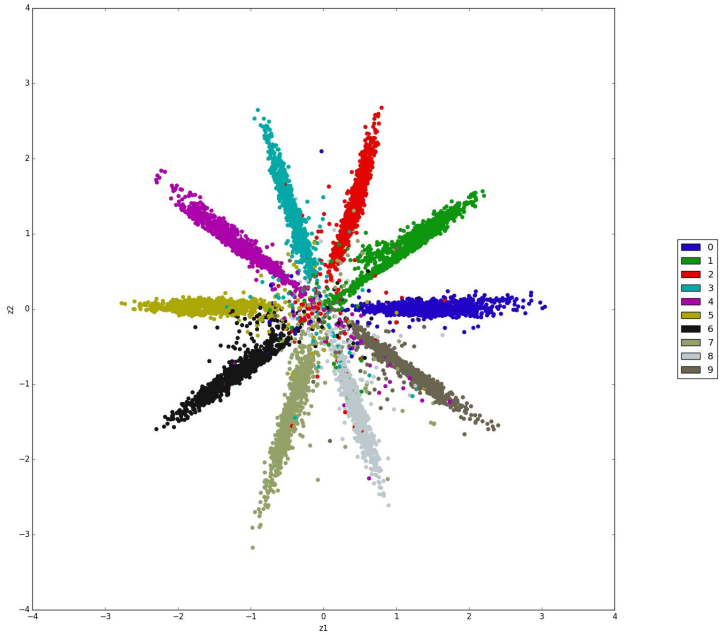
Exercise

- Create linear autoencoder
- Print 2d representation of dataset
- Create non-linear autoencoder and compare

Linear Autoencoder

Tutorial (link)

Exercise (link)


4
3
2
1
0
−1
−2
−3
−4
z2
−4
−3
−2
−1
0
1
2
3
4
z1
0
1
2
3
4
5
6
7
8
9


Input
VAE
$\text{VAE}_{\text{Dis}_l}$
VAE/GAN

Stacked Autoencoder (with activation)

Exercise (link)

Homework - Fashion Mnist

Homework (link)