Autoencoders and Denoising Autoencoders
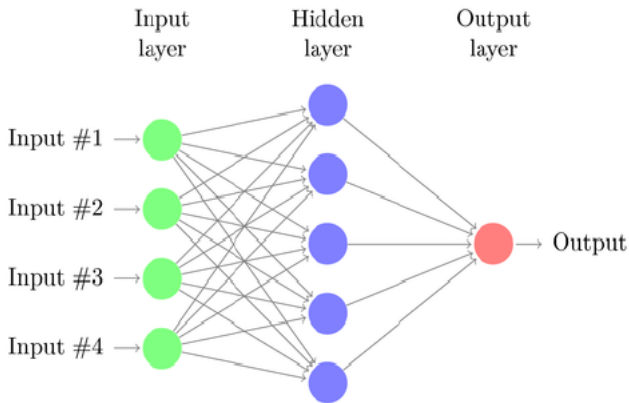
Lovedeep Gondara

MLRG, March 2018

Multilayer NN

Figure: Generic Neural Network

Multilayer NN

- We have labelled training examples
- Train the network to minimize output error (predicted vs real)
- i.e. Supervised training

Autoencoder

- We don't have training labels, just a dataset, unsupervised?
- What if we use the input as the labels
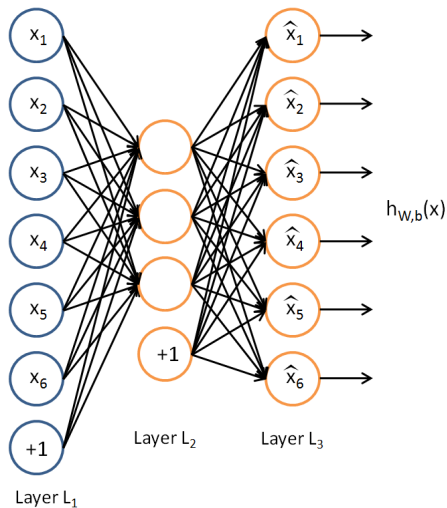- Network tries to reconstruct the input, self-supervised?

Autoencoder

Figure: A simple autoencoder

Autoencoder

- Trying to learn an approximation to the identity function
- Identity function: $f(x) = x$
- Seems trivial, why do we care?

Autoencoder

- We can learn some interesting things from intermediate representations
- What if the number units in hidden layer are smaller than input?
- What if the number of units in hidden layer are larger than input?

Autoencoder

- We can learn some interesting things from intermediate representations
- What if the number units in hidden layer are smaller than input?
- What if the number of units in hidden layer are larger than input?

Autoencoder

- Can stack individual autoencoders to create a stacked deep autoencoder
- Yield better insights for large data
- Layerwise pre-training

Autoencoder

- Application to NLP: Word embeddings, Machine translation, Document clustering etc.
- Generic: Dimensionality reduction, Compression, Anomaly detection, pretraining etc.
- Image: Inpainting

Denoising Autoencoder

- ▶ Same as a generic autoencoder
- ▶ Input is corrupted
- ▶ Reconstruction is based on non-corrupted data
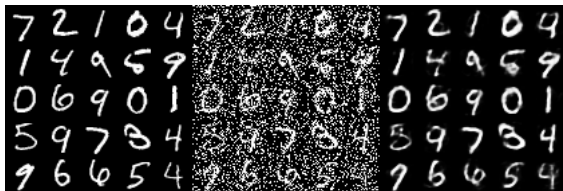
Denoising Autoencoder

Figure: Denoising example

Denoising Autoencoder

- Force the hidden layer to learn more robust features about the input while dealing with the noise/corruption
- i.e. not just an identity function
- Denoising autoencoder is a stochastic version of autoencoder

Denoising Autoencoder

- Tries to simultaneously encode the input and remove noise
- Noise is added stochastically at the input layer
- Different angles to understand the working of denoising autoencoder (manifold perspective, information theoretic, generative etc.)
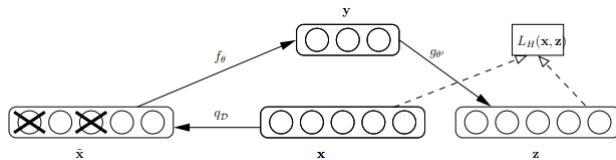
Denoising Autoencoder

Figure: Denoising autoencoder

Denoising Autoencoder

- Robust representation: Learns from multiple input data representations
- To predict any subset of variables from the rest is a sufficient condition for completely capturing the joint distribution between a set of variables

Denoising Autoencoder

- Application to imaging: Denoising/completion
- Also can do most of the stuff an autoencoder can

Autoencoder/denoising autoencoder

- Generic design
- Can be constructed with any type of NN (CNN, RNN etc.)

References I

📄 Le, Quoc V
A tutorial on deep learning part 2: autoencoders, convolutional neural networks and recurrent neural networks.
Google Brain, (2015): 1-20.

📄 Goodfellow, Bengio and Courville
Deep learning book, Chapter 14: Autoencoders
MIT press, (2016)

📄 Stanford deep learning course
http://ufldl.stanford.edu/tutorial/unsupervised/
Autoencoders/