Introduction to Neural Networks & Deep Learning

S. R. M. Prasanna

Dean (Faculty Welfare, Research & Development)
Professor, Dept of Electrical Engineering
Indian Institute of Technology Dharwad

prasanna@iitdh.ac.in

July 21, 2021

Some Good Resources for NNDL

- Duda and Hart, "Pattern Classification" Wiley, 2001.
- C. M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.
- B. Yegnanarayana, "Artificial Neural Networks", PHI, 1999.
- Michael Nielsen "Neural Networks and Deep Learning" Open Book.
- Ian Goodfellow, Yoshua Bengio and Aaron Courville, "Deep Learning", MIT Press, 2016.
- Charu Agarwal, "Neural Networks and Deep Learning", Springer 2018.
- Mithesh M. Khapra, "Deep Learning Part I and II" NPTEL Course.
- Ali Ghodsi, "Deep Learning", University of Waterloo Course.

Outline

- Introduction
- Neural networks & its shortcomings
- Deep learning & its training
- Feature extraction or representation learning by DL
- Compression by DL
- Distribution modelling by DL
- Classification modelling by DL
- Time series modelling by DL
- Summary

Neural Networks

- Inspired by the structure and functioning of biological neural networks
- Parallel and distributed processing
- Instead of one central processor, a large no of computing elements
- Power for NN comes from parallel and distributed network
- NN needs different way of data processing
- Conventional, one entry for data, sequential processing and one exit
- NN, parallel entry, parallel processing and parallel exit
- NN: Redundancy, graceful degradation, nonlinear processing ...

Programming: Central Processor vs NN

- Central: Flowchart / algorithm, program, sequential
- NN: Flowchart / algorithm, program, parallel. Learns from data and figuring out its own solution.
- Automatically learning from data sounds promising $\Longrightarrow$ motivation for machine learning (ML) and deep learning (DL)
- Until 2006 didn't know how to train neural networks (Deep Networks) that have many layers to surpass traditional approaches.
- Vanishing or exploading gradient problem
- Demonstration of greedy layer wise training in 2006.
- Deep learning seem to have achieved human performance level in computer vision and speech processing.

Deep Learning Motivation: Human way of Learning

- Learning from experience
- Need not specify everything in the beginning
- Understand in terms of hierarchy of concepts
- Each concept defined in terms its relation to simpler concepts
- Learning complicated concepts out of simpler ones

Distinction Among Terminologies

- Pattern recognition: Hand crafted features + learning pattern as a mathematical model.
- Machine learning: Hand crafted features + learning pattern as a function approximation (data driven).
- Deep learning: Representation learning by machine (data driven) + pattern discovery as a layered approach (data driven).

Neural Networks and Deep Learning

- Genesis of ML is in PR.
- Genesis of DL is in NN.
- Like PR is basis for ML.
- NN is basis for DL.
- DL enables both learning features and also modeling information in a better manner.
- Depth is not only in terms of no of layers, but also in terms of no of learning functions (concepts) and relating different concepts.
- DL $\in$ ML $\in$ AI.

History and Breakthroughs of DL

- Geoffrey Hinton's group $\Longrightarrow$ a deep belief network could be efficiently trained using a greedy layer-wise pre-training (2006)
- Other research groups, same strategy be used to train many other kinds of deep networks (2007)
- Able to train deeper neural networks than had not been possible.
- DNNs outperformed AI systems based on other machine learning technologies & hand crafted features.
- The models we train have undergone changes that simplify the training of very deep architectures.

Fuelling Growth of DL

- GPUs, Memory
- Software infra, open source toolkits, distributed computing
- Codes for different deep learning models in public domain
- Large amount of data

Artificial Neural Networks for Acoustic Modelling

- ANNs trained using BP have potential to learn better models of data that lie on or near a non-linear manifold.
- Bourlard and Morgan shown ANNs with a single layer of non-linear hidden units to predict HMM states.
- Neither hardware nor learning strategy were adequate for training neural networks with many hidden layers on large amounts of data.
- Performance benefits of using ANNs with a single hidden layer or 2-3 layers were not sufficiently large to seriously challenge GMM-HMM.
- Main contribution of neural networks at that time was to provide extra features in tandem or bottleneck systems.

H. Bourlard and N. Morgan, Connectionist Speech Recognition: A Hybrid Approach, Kluwer Academic Publishers, Norwell, MA, USA, 1993.

G. Hinton et. al, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups", Signal Processing Magazine, vol. 29 (6), pp. 82-91, Aug. 2012.

Issues in Training DNN

- DNN's with many hidden layers are hard to optimize.
- Backpropagated gradients will have very different magnitudes (vanishing or exploding gradient) in different layers.
- Very large training sets can reduce overfitting thus preserving modeling power, but training computationally expensive.
- What we need is a better method of modelling information in training data to build multiple layers of non-linear feature detectors.

G. Hinton et. al, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups", Signal Processing Magazine, vol. 29 (6), pp. 82-91, Aug. 2012.

Two stage training procedure for DNN

- In the first stage, layers of feature detectors are initialized, one layer at a time, by fitting a stack of generative models, each of which has one layer of latent (hidden) variables.

- In the second stage, each generative model in the stack is used to initialize one layer of hidden units in a DNN and the whole network is then discriminatively fine-tuned to predict the target HMM states.

G. Hinton et. al, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups", Signal Processing Magazine, vol. 29 (6), pp. 82-91, Aug. 2012.

Generative Pre-Training of DNN

- Instead of designing feature detectors to be good for discriminating between classes, design them at modeling the structure in input data.
- Learn one layer of feature detectors at a time, with states of feature detectors in one layer acting as data for training next layer.
- After generative "pre-training", multiple layers of feature detectors can be used as a much better starting point for a discriminative "fine-tuning" during which backpropagation through the DNN slightly adjusts the weights found in pre-training.
- The generative pre-training finds a region of the weight-space that allows the discriminative fine-tuning to make rapid progress, and it also significantly reduces overfitting.

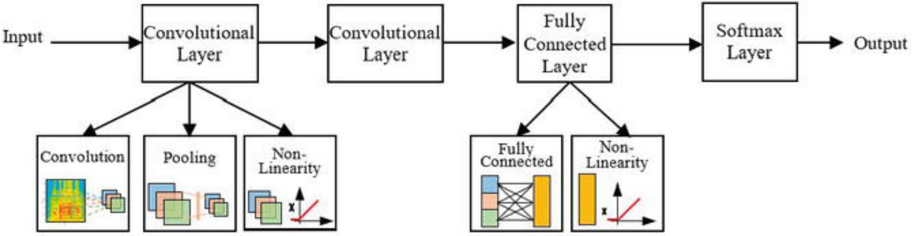
G. Hinton et. al, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups", Signal Processing Magazine, vol. 29 (6), pp. 82-91, Aug. 2012.

Performance on TIMIT Database

- GMM-HMM gave a phoneme error rate (PER) of 27.3%.
- Initial ver of DNN using randomly initialized weights gave 23.4%.
- Monophone DBN-DNNs gave 22.4%.
- Monophone convolutional DNNs on fbank gave 20%.

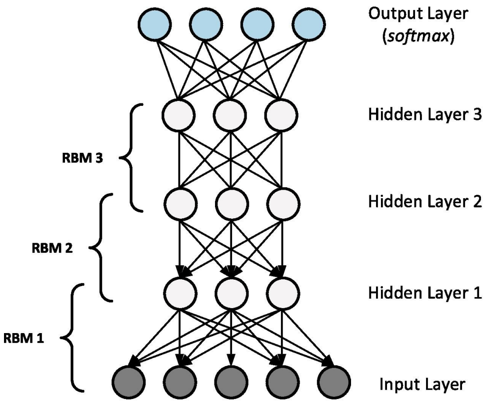
G. Hinton et. al, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups", Signal Processing Magazine, vol. 29 (6), pp. 82-91, Aug. 2012.

Feature Extraction using Deep Learning: Convolutional Neural Netowrk (CNN)

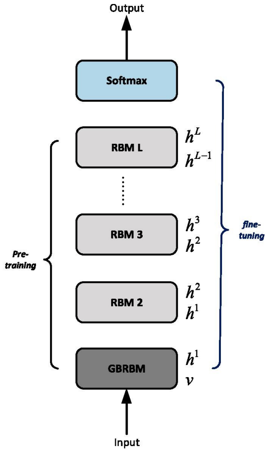


[taken from public domain and original authors are highly acknowledged]

Feature Extraction using Deep Learning: Deep Belief Network (DBN)



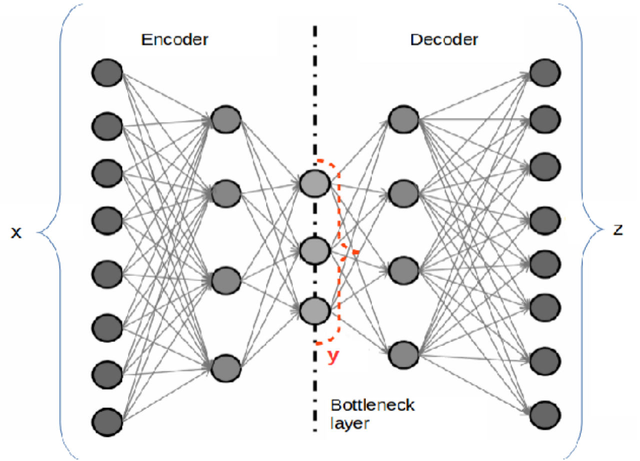
(a)



(b)

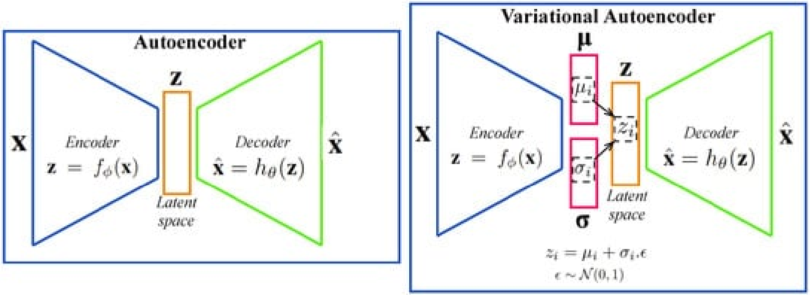
[taken from public domain and original authors are highly acknowledged]

Feature Compression using Deep Learning: Auto Encoder



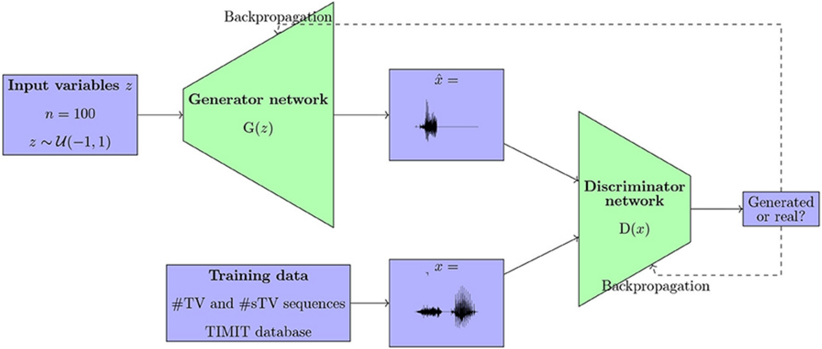
[taken from public domain and original authors are highly acknowledged]

Distribution Modeling using Deep Learning: Variational Auto Encoder



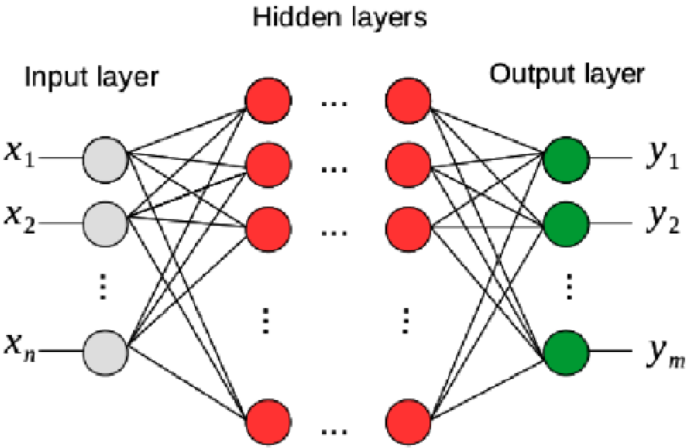
[taken from public domain and original authors are highly acknowledged]

Distribution Modeling using Deep Learning: Generative Adversarial Network



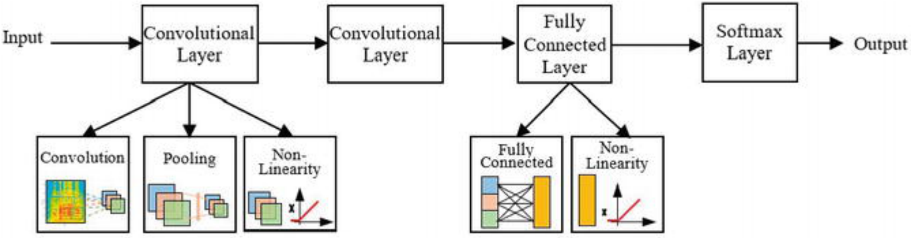
[taken from public domain and original authors are highly acknowledged]

Classification using Deep Learning: Deep Forward Neural Network (DFNN)



[taken from public domain and original authors are highly acknowledged]

Classification using Deep Learning: CNN



[taken from public domain and original authors are highly acknowledged]

Time series modelling using Deep Learning

- Time delay neural network (TDNN)
- Recurrent neural network (RNN)
- Long-shorterm memory (LSTM)

Summary

- Introduction to neural networks
- Breakthrough in deep learning
- Training deep learning networks
- Feature extraction using DL
- Compression using DL
- Distribution modeling using DL
- Classification using DL
- Time series modeling uisng DL

Thank You