Representation Learning and Deep Neural Networks

Davide Bacciu

Dipartimento di Informatica
Università di Pisa
bacciu@di.unipi.it

Applied Brain Science - Computational Neuroscience (CNS)

Learning Representations in the Brain

- Sensory information is represented by neural activity
  - Response selectivity of individual neurons
  - Distribution of activation in neural population
  - Something we have seen so far
- Neural representation is hierarchical
  - Brain cortex processes information incrementally
  - Increasing levels of abstraction

Representation Learning

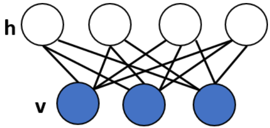
How can we obtain articulated hierarchical representations of information in computational models?

Representation Learning - A Computational View

Learning the unknown causes underlying data

- Causes explain the data we observe
- Are also the basic building blocks which, when combined, generate the data we observe

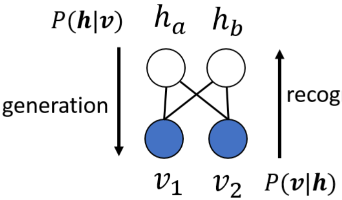
Something we have already seen with RBM

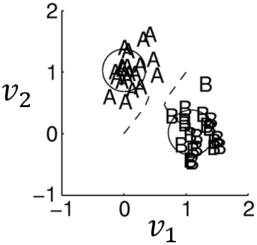


- Try explaining input data **v** using the unknown causes **h**
- Learning is
  - Generative as causes can reconstruct the data
  - Unsupervised as interest is in representing information rather than, e.g., predicting a class

Representation Learning - A Classical View

Representation learning as density estimation: learn a probability distribution for the data **v** that uses latent variables **h**





Learning of a Gaussian Mixture Model

- Data likelihood $P(\mathbf{v}|\mathbf{h})$
- Posterior $P(\mathbf{h}|\mathbf{v})$

Several models in this classical view: PCA, ICA, factor analysis, clustering, ...

Representation Learning - A Modern View

Learn representations (a.k.a. causes or features) that are

- Hierarchical: representations at increasing levels of abstraction
- Distributed: information is encoded by a multiplicity of causes
- Shared among tasks
- Sparse: enforcing neural selectivity
- Characterized by simple dependencies: a simple (linear) combination of the representations should be sufficient to generate data

Deep models follow this modern view of representation learning

What is Deep Learning?

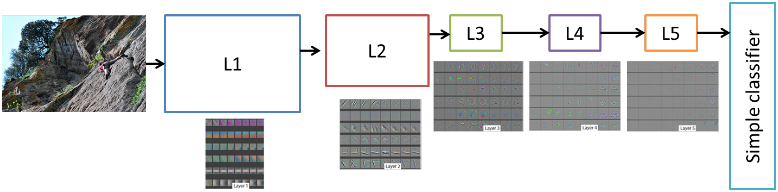
Machine learning algorithms inspired by brain organization, based on learning multiple levels of representation and abstraction

- Learning models with many layers trained layer-wise
- Build a hierarchical feature space through layering
- Reduce the need of supervised information
  - Unsupervised discovery of features in the internal layers
  - Final layer performs supervised step

Learning Features

The traditional shallow way

The deep way



Why Deep Learning?

Google

DEEPMIND

DNNresearch

facebook

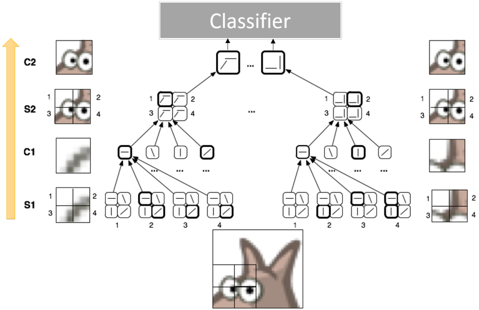
No, Seriously.. Why Deep Learning?

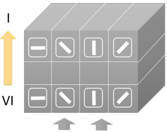
- Deep learning is THE hot topic now
- Revolutionized performance in
  - Speech recognition
  - Machine vision
  - Natural language processing
- Now expanding to other topics
  - Reinforcement learning
  - Robotics
  - Bioinformatics and Cheminformatics

Its origins and inspiration can be traced back to brain/cortex organization

Hierarchical Representation
HMAX Model

Explaining the structure of the visual cortex in mammals



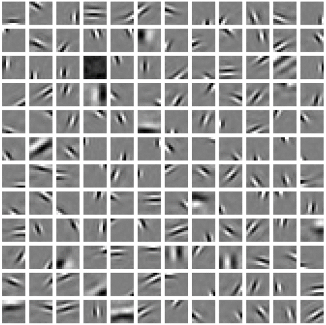


M. Riesenhuber, T. Poggio, Hierarchical Models of Object Recognition in Cortex. Nature Neuroscience 2: 1019-1025, 1999

Sparse Representation

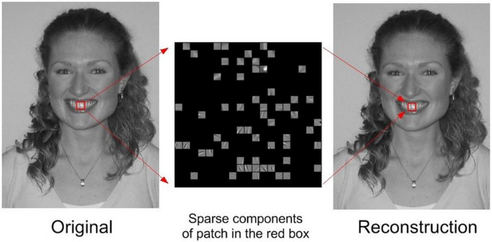
Neuroscience

Sparse coding theory in brain: sensory information in the brain is represented by a relatively small number of simultaneously active neurons out of a large population (B.A. Olshausen, D.J. Field, 1996)



Sparse Representation
Machine Learning

The machine learning perspective: a single layer network learns better to generate a target output if the input has a sparse representation (Willshaw and Dayan, 1990)



Deep Networks

- Deep architecture with multiple layers focused on learning sparse encoding of input data
- Unsupervised training between layers to decompose the problem into distributed sub problems with increasing levels of abstraction
- Deep networks type
  - Deep Generative Models
  - Stacked Neural Autoencoders
  - Deep convolutional neural networks
  - Recurrent neural networks
  - Hybrids of the above

Training Deep Networks

Problem with conventional backpropagation training

- Strongly relies on labeled training data
- Learning does not scale well to multiple hidden layers

Greedy layer-wise training

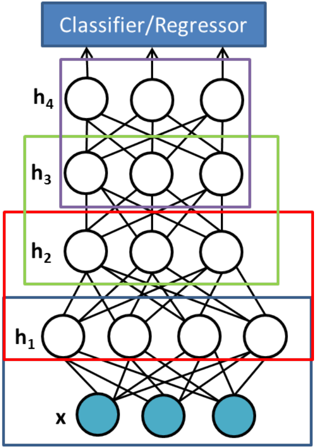
- Unsupervised (internal) layer by layer representation learning (pre-training)
- Supervised training of last layer (read-out) plus optional fine-tuning

Key advantages

- Give full learning focus to each layer
- Exploit unlabeled data using supervised training only for fine tuning

Deep Generative Models

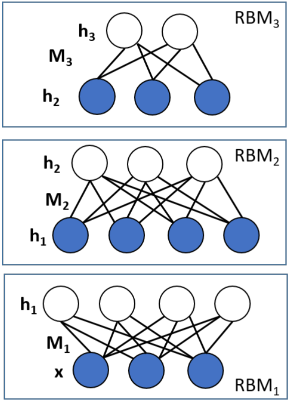
Intuition - Train a network where hidden units **h** organized into a hierarchy extract a good representation data from visible units **x**

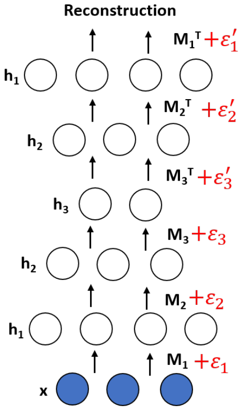


Architecture - A network of stacked Restricted Boltzmann Machines (RBM) plus a supervised read-out layer for making predictions

Deep Belief Networks (DBN)

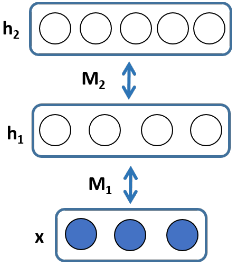
Pretraining of the independent RBM by Contrastive-Divergence, which are then unrolled to form the deep architecture that is fine-tuned by error backpropagation

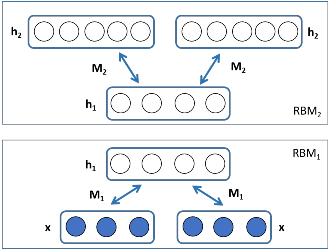




Deep (Restricted) Boltzmann Machines

DBM are directed generative models ⇒ To train a Deep RBM we need a pre-training trick





Sampling $\mathbf{h}_1$ by averaging bottom-up and top-down contribution

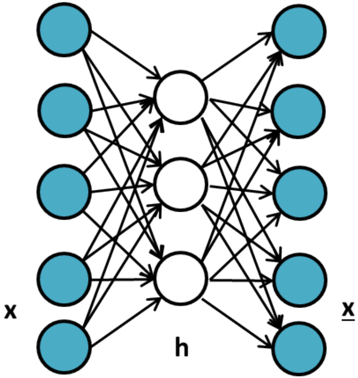
$$h_{1j} \sim \sigma\left(\sum_i M_{ij} x_i + \sum_m M_{jm} h_{2m}\right)$$

For multiple layers

- Apply modified training to first and last layer
- Halve the weights of inner layers

Neural Autoencoder Networks

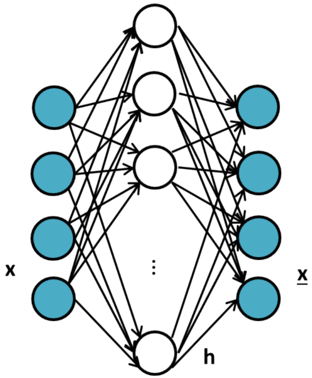
Non-accretive associative memories for feature discovery and data compression



E.g. linear hidden units with MSE perform PCA (guess why?)

Sparse Autoencoders

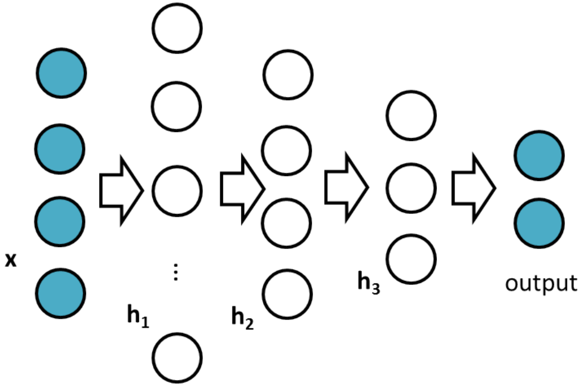
Autoencoders using more features with a significant number of neuron being 0-active when encoding an input



Using regularization approaches to enforce sparsity

Stacked Neural Autoencoders

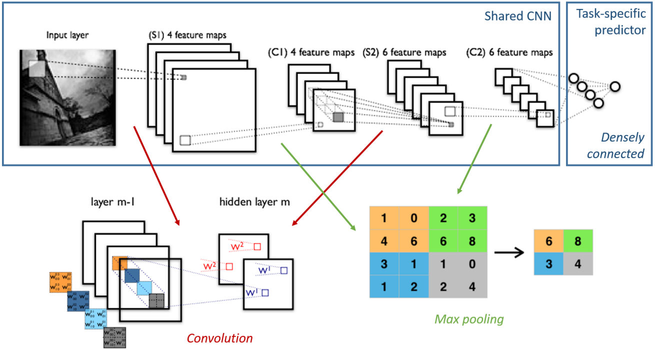
Stack autoencoders with level-wise training as with DBM



Train each autoencoder in isolation and drop the decoding layer when training is completed

Convolutional Neural Networks (CNNs) - LeNet

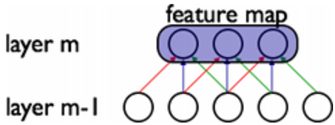
Very much inspired by the visual processing pipeline in the brain



Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86(11): 2278-2324, 1998

Strategies to Control Model Complexity (Regularization)

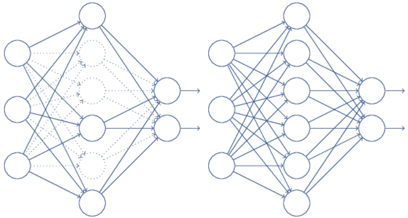
Weight Sharing



- Exploit symmetry or stationarity assumption to reduce the number of parameters
- Convolutional NN, recurrent and recursive NN

Dropout

- Randomly disconnect hidden neurons during training
- Need to scale hidden weights at test time



Stochastic Gradient Descent

Mini-batches

- Compute stochastic gradient on small batches of data rather than on single sample
- Stabler and quicker learning and facilitates GPU computing as a byproduct

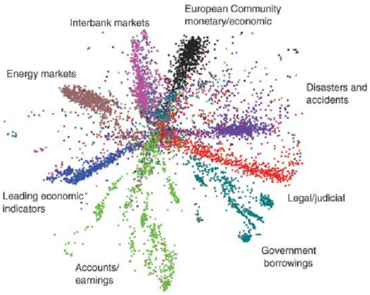
Momentum Method

$$\Delta w(t+1) = \alpha \Delta w(t) - \epsilon \frac{\partial E}{\partial w}(t+1)$$

- Changes the velocity of the weight particle instead of the position
- May quicken convergence due to avoiding oscillations

Document Encoding - Deep Belief Network

Finding sparse features for $> 800K$ newswire stories



DBN document encoding

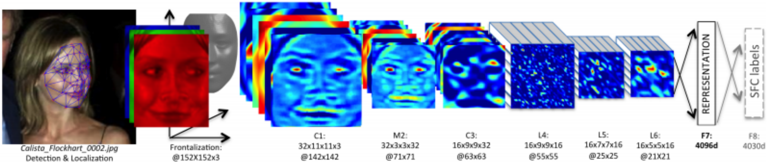
Latent semantic analysis

Hinton, Salakhutdinov, Reducing the dimensionality of data with neural networks, Science, 2006

DeepFace: a.k.a. beating humans at recognizing faces

DeepFace recognition accuracy is $\approx 97.35\%$

- 23% better than previous results
- CNN with more than 120 million parameters

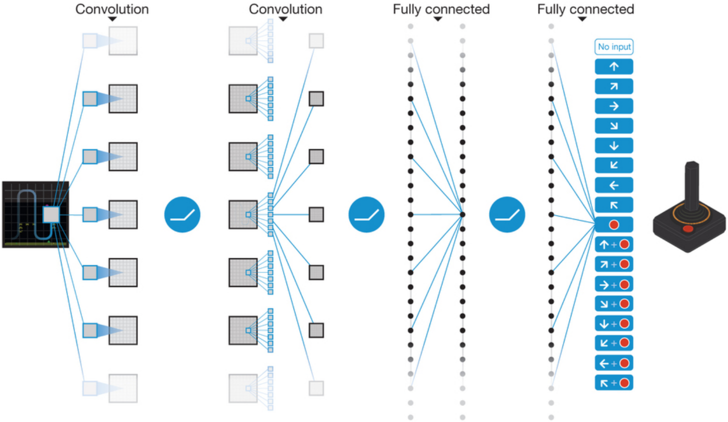


facebook

Taigman et al, Deepface: Closing the gap to human-level performance in face verification, CVPR 2014

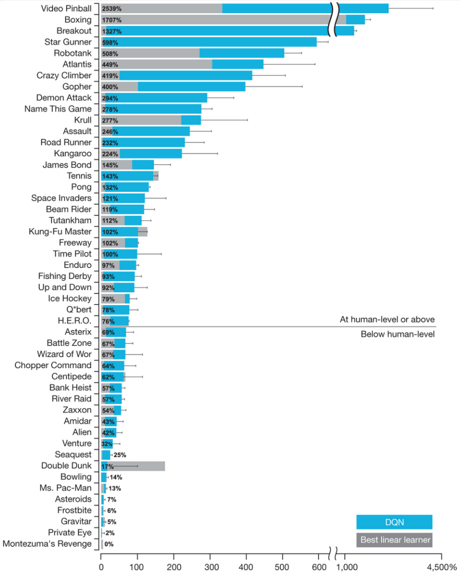
Learning to Play Atari (I)

Integrating CNN with Reinforcement Learning



Mnih et al, Human-level control through deep reinforcement learning, Nature 2015

Learning to Play Atari (II)



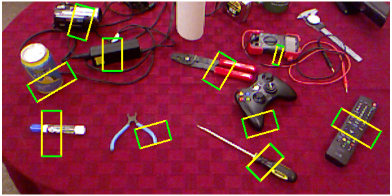
Deep Learning and Robotics

Learning where to grasp objects with robotic manipulators

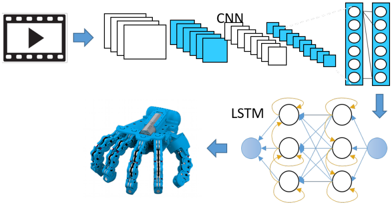
- Deep auto-encoder network
- Predicts which grasping box works better

Learning how to approach and grasp objects from humans

- A combination of CNN and deep recurrent nets
- Train on few controlled video to analyse uncontrolled Youtube clips



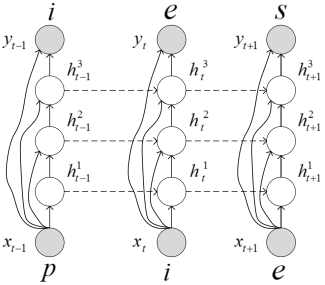
Lenz et al, Deep Learning for Detecting Robotic Grasps, IJRR 2014

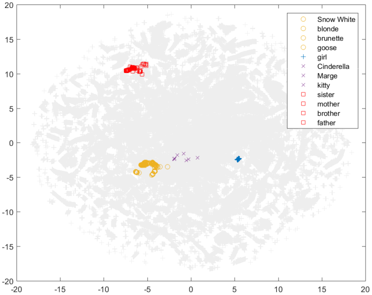


Work in progress with the Softhand group at Centro Piaggio

You can also have FUN (2016) with DL...

Learning to generate English jokes char-by-char

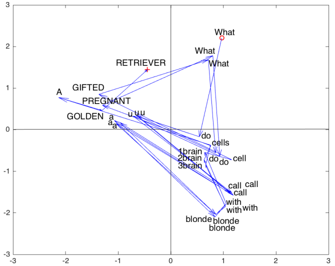




Bacciu, D., Gervasi, V., Prencipe, G. LOL: An Investigation into Cybernetic Humor, or: Can Machines Laugh?. FUN 2016

You can also have FUN (2016) with DL...

You can trace joke climax in the neurons...



What do u call a blonde with 1 brain cell? GIFTED!
What do u call a blonde with 2 brain cells? PREGNANT!
What do u call a blonde with 3 brain cells? A GOLDEN RETRIEVER!

..and you can of course generate (poor) jokes:

What do you call a car that feels married? A cat that is a beer!

What do you get if you cross a famous california little boy with an elephant for players? Market holes.

Why did the boy stop his homework? Because they're a bunny boo!

Deep Learning API

- Caffe - DL framework for computer vision
  - http://demo.caffe.berkeleyvision.org/
- Theano - Python library with GPU support
  - http://deeplearning.net/software/theano/
- Torch - Scientific computing environment
  - http://torch.ch/
- MatConvNet - CNN in Matlab with GPU support
  - http://www.vlfeat.org/matconvnet/
- CNTK - Microsoft NN and DL toolkit
  - http://www.cntk.ai/
- Tensorflow - Google NN and DL library
  - https://www.tensorflow.org/

Make sure you have a good GPU!

Want to try something easy out?

- Google Tensorflow Playground
  - `http://playground.tensorflow.org`
- Language learning and generation
  - `http://karpathy.github.io/2015/05/21/rnn-effectiveness/`
- Caffe image classification demo
  - `http://demo.caffe.berkeleyvision.org/`
- Nvidia Digits deep learning GUI
  - `https://developer.nvidia.com/digits`
- Intel deep learning training tool
  - `https://software.intel.com/en-us/deep-learning-training-tool`

Where is DL heading?

- Image and video understanding by integration with natural language descriptions
  - Bringing in biological models of attention
- Natural language understanding and generation
  - Emotion understanding
  - Moving from sequential representation of text to parse trees
- Is starting to impact heavily in robotics
  - Sensorimotor learning
  - Policy learning: deep reinforcement learning
  - Cloud robotics: deep learned knowledge available to connected devices
  - Onboard GPU computing
  - Self-driving cars, drones, connected devices

Take Home Messages

- Deep learning is about learning features of complex data
  - Features $\equiv$ hidden stochastic neurons
  - Strongly rooted in hierarchical information processing in the brain
- Stacking and level-wise training
  - Let the deep network discover the features and then place your preferred learning model to perform your task
- Breakthrough performance in several learning tasks/application areas
  - Complex spatio/temporal structured data
- Do we really know what is going on with the encoding? How many layers?

Next Lecture

Will be on May the 3rd

- Topics for final presentations and coding projects
- Last lab assignment
  - Implementing restricted Boltzmann machines