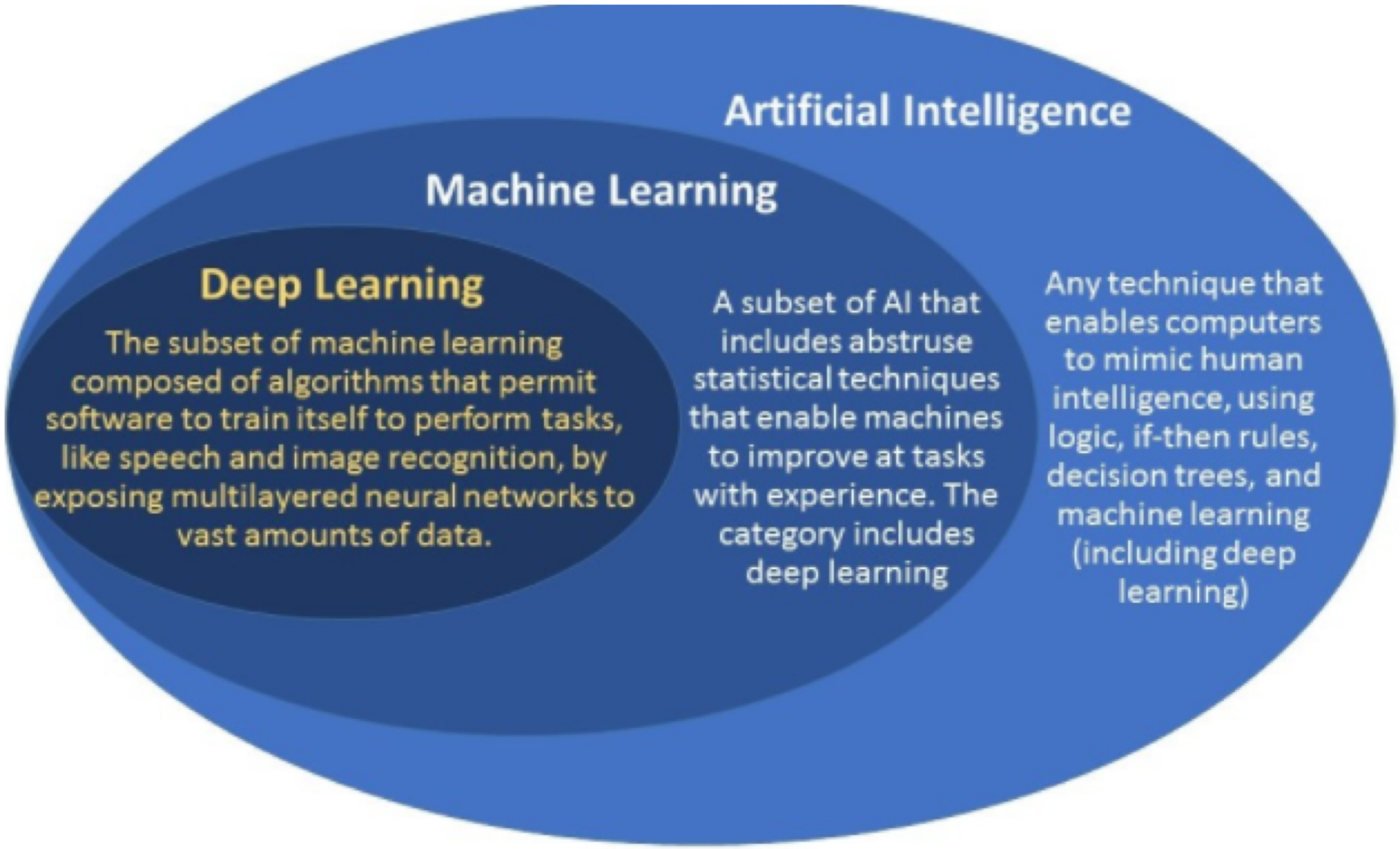
Computing Methods for Experimental Physics and Data Analysis

Data Analysis in Medical Physics

Lecture 10: Machine learning and deep learning

Alessandra Retico
alessandra.retico@pi.infn.it
INFN - Pisa

Artificial Intelligence, Machine Learning, Deep Learning



https://www.deeplearningitalia.com/una-panoramica-introduttiva-su-deep-learning-e-machine-learning/

Machine Learning

- **Supervised learning:**
  - Predicting values, **known** data labels
  - The machine uses the label information to guess the right answer on new data

Regression
Estimate continuous values
(real-valued output)

Classification
Identify a unique class
(Boolean, discrete values, categories)

- **Unsupervised learning:**
  - Search for structure in data; **unknown** data labels
  - The machine find useful information hidden in data

Cluster Analysis
Group data points into sets

Dimensionality reduction
Select relevant variables

Regression predictive models

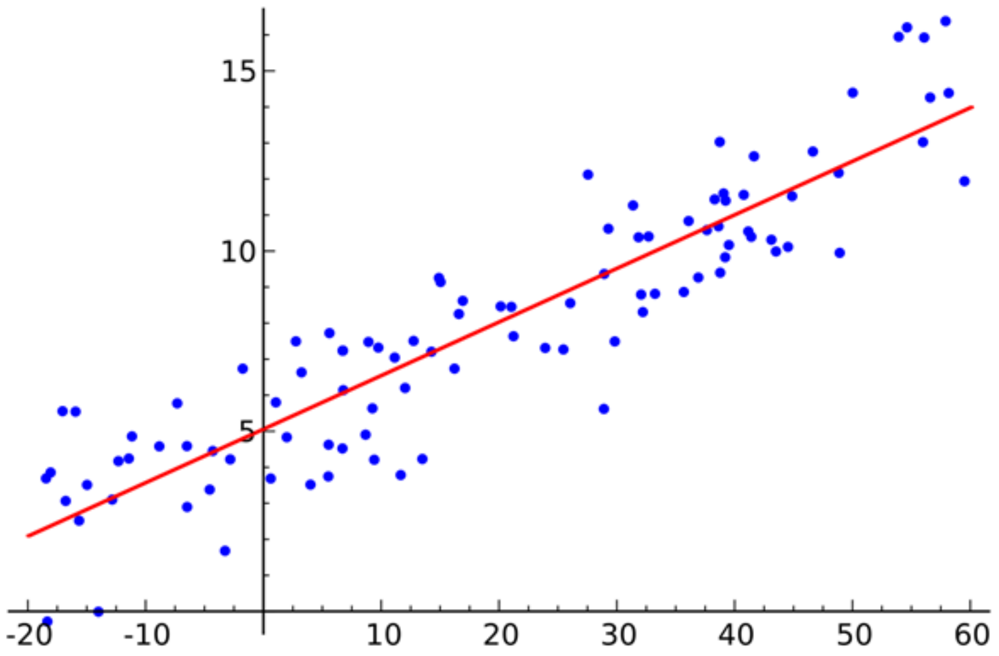
- Supervised learning: it aims to model the relationship between a certain number of features (multivariate regression problem) and a continuous target variable.

$$y = f(x)$$

- the most common figure of merit used to estimate the performance of a regression predictive model is the root mean squared error (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

Simple Regression → $y = mx + b$



See demo code on
https://github.com/retico/cmepda_medphys
L10_code/Lecture10_demo_regression.m

Autoencoders

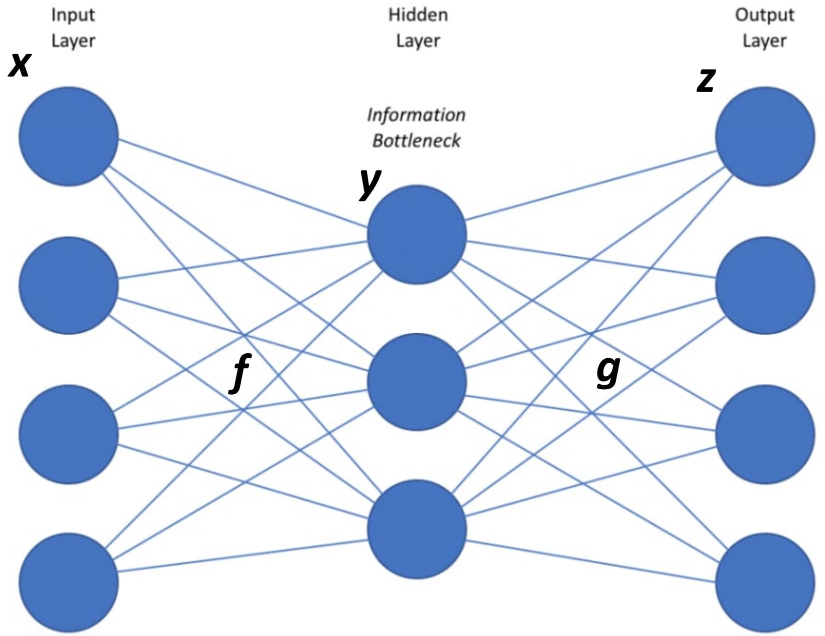
Unsupervised learning.

Artificial Neural Network able to compress (encode) information automatically.

An *encoder* is a deterministic mapping ***f*** that transforms an input vector ***x*** into hidden representation ***y***

A *decoder* maps back via ***g*** the hidden representation ***y*** to the reconstructed input ***z***.

An **autoencoder** compares the reconstructed input ***z*** to the original input ***x*** and try to minimize the reconstruction *error*.



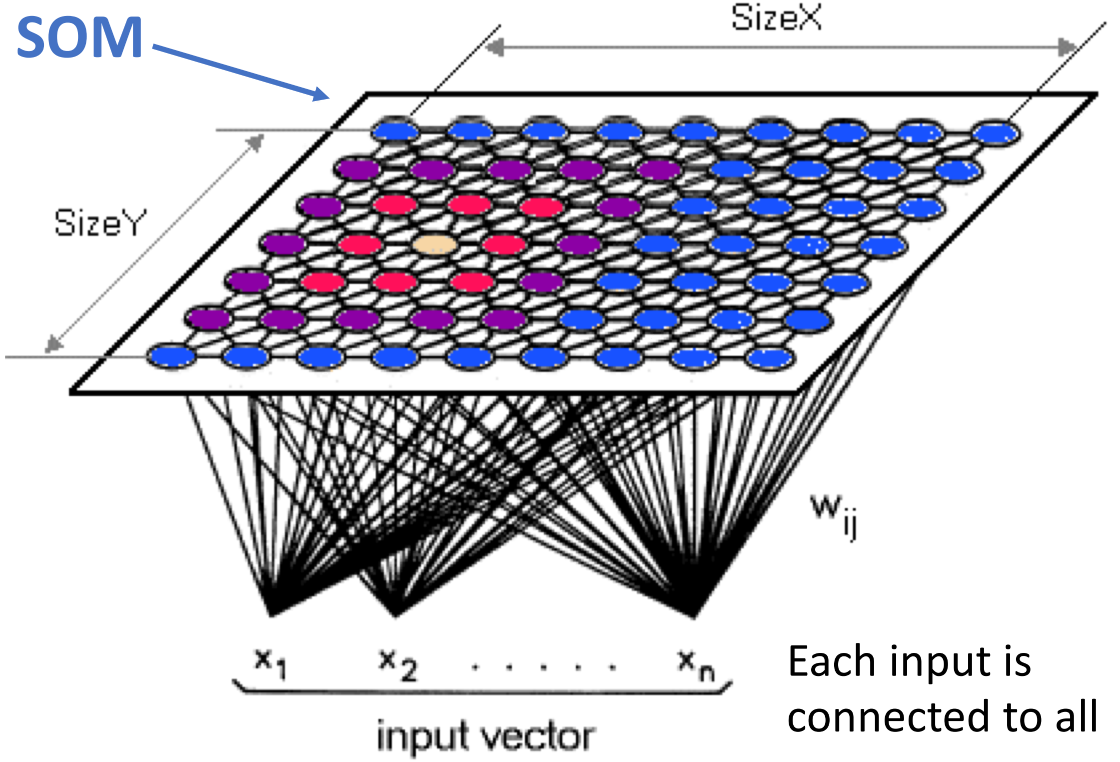
Autoencoders can be used:

- to compress the information (using the higher level representations ***y***).
- to denoise images, as ***y*** are relatively stable and robust to input corruption.

Bourlard, H.; Kamp, Y. (1988). "Auto-association by multilayer perceptrons and singular value decomposition". *Biological Cybernetics*. **59** (4–5): 291–294

Self Organizing Maps (SOM)

- Unsupervised learning.
- Unsupervised Artificial Neural Networks that maps multidimensional data onto a 2 dimensional grid
- SOM are known also as Kohonen feature maps, as they were introduced by T. Kohonen in 1982
- SOM combine a **competitive learning** principle with a topological structuring of neurons such that adjacent neurons tend to have similar weight vectors.

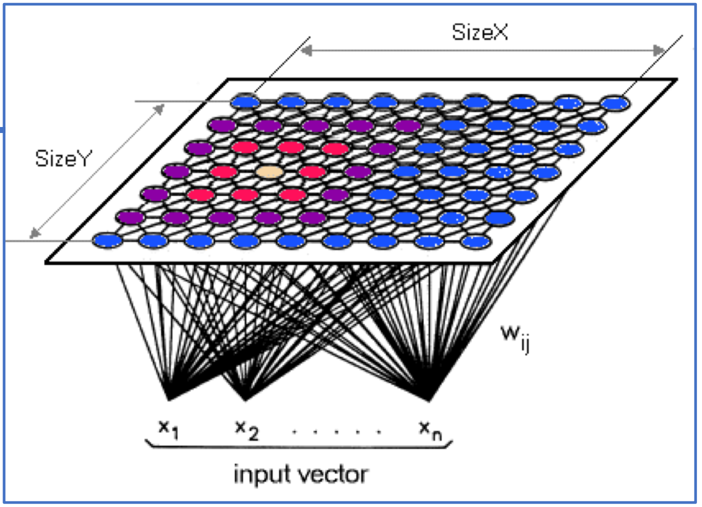


Each input is connected to all output neurons

- SOM can be used for detecting similarity and degrees of similarity
- Underlying neurobiological hypothesis:
    - Structure self-organises based on learning rules and system interaction.
    - Axons physically maintain neighborhood relationships as they grow.

Teuvo Kohonen. Self-Organization of very large document collections: State of the art (1998)

Self Organizing Maps (SOM)



Define the SOM size: choose the size of the SOM (e.g. NxN neurons)

Initialization: choose random small values for weights such that

$w_j(0)$ is different for all neurons j.

Sampling: get a sample example x from the input space.

Similarity matching: find the best matching (Euclidean dist.) winning neuron $i(x)$ at step n:

$$i(x) = \arg\min_j \| x(n) - w_j \| \qquad j \in [1,2, \ldots, NxN] \qquad \rightarrow i \text{ is the winner neuron}$$

Updating: adjust synaptic weight vectors of winning neuron i and its neighbors

$$w_j(n+1) = w_j(n) + \eta(n)\, h(j,i,n)\, (x - w_j(n))$$

where η(t) is a leaning coefficient (decreasing with increasing n) and $h(j,i,n)$ defines the neuron neighboring condition, which has its max for $j=i$ (it can be Gaussian shaped) and its extent can decrease with increasing n.

Continuation: go to Sampling step until no noticeable changes in the feature map are observed.

SOM interpretation

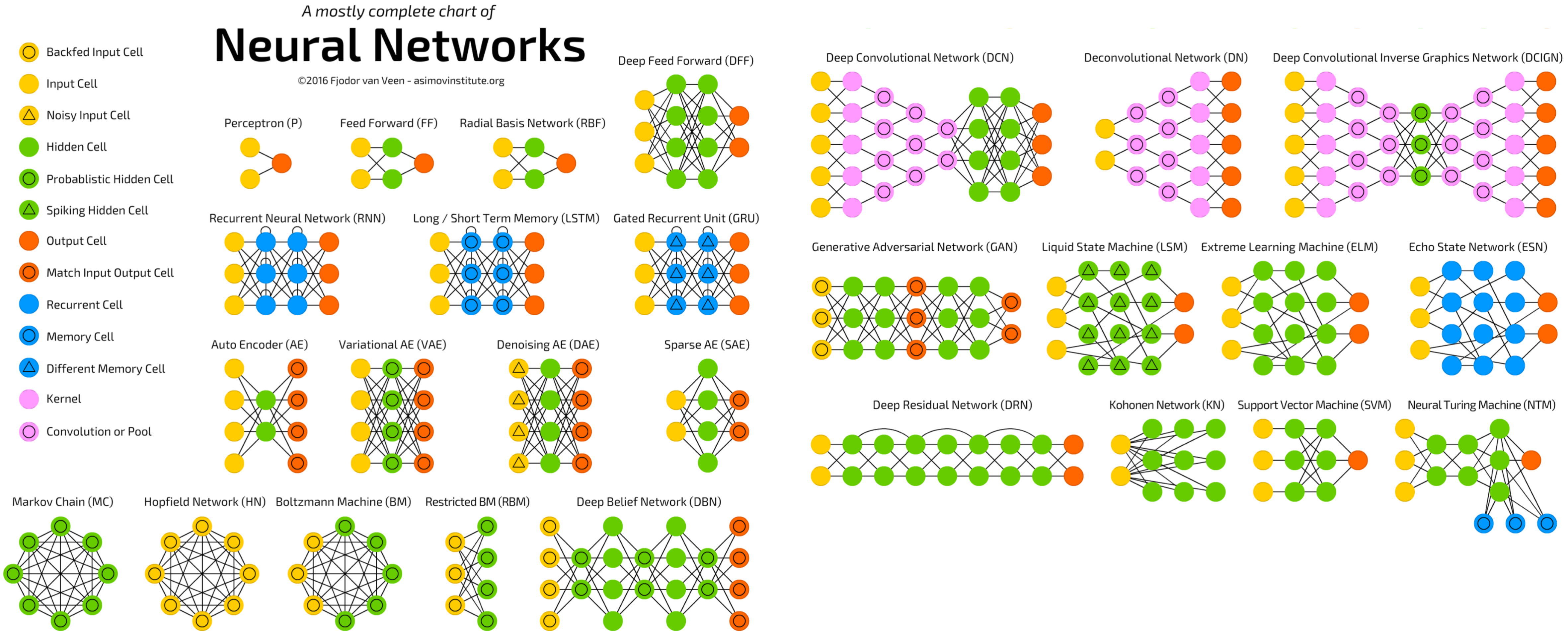
- There are two ways to interpret a SOM.
  - Because in the training phase weights of the whole neighborhood are moved in the same direction, similar items tend to excite adjacent neurons. Therefore, SOM forms a semantic map where similar samples are mapped close together and dissimilar ones apart.
  - The other way is to think of neuronal weights as pointers to the input space. They form a discrete approximation of the distribution of training samples. More neurons point to regions with high training sample concentration and fewer where the samples are scarce.

Data-mining applications: discovering similarities in data.

Large SOMs display emergent properties. In maps consisting of thousands of nodes, it is possible to perform cluster operations on the map itself.

See demo code on https://github.com/retico/cmepda_medphys
L10_code/Lecture10_demo_SOM.m

Neural network Zoo

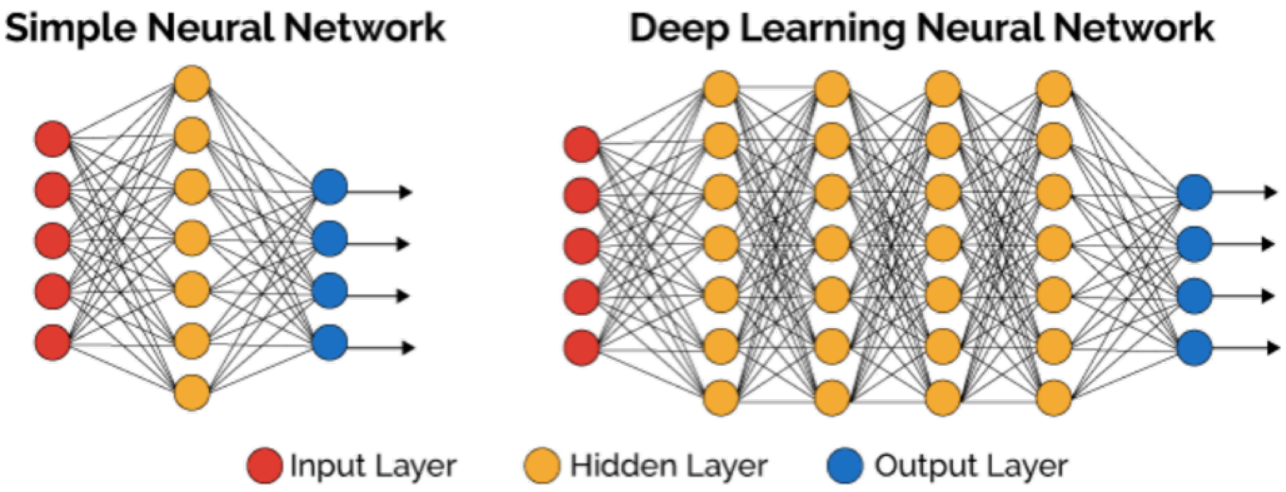


http://www.asimovinstitute.org/neural-network-zoo/

Deep Neural Networks

Deep Learning (DL) means using a neural network with several layers of nodes between input and output

DL models are a family of parametric models which learn non-linear hierarchical representations



$$a_L(\mathbf{x};\boldsymbol{\Theta}) = h_L(h_{L-1}(\ldots(h_1(\mathbf{x},\boldsymbol{\theta}_1),\boldsymbol{\theta}_{L-1}),\boldsymbol{\theta}_L)$$

input

parameters of the network

non-linear activation function

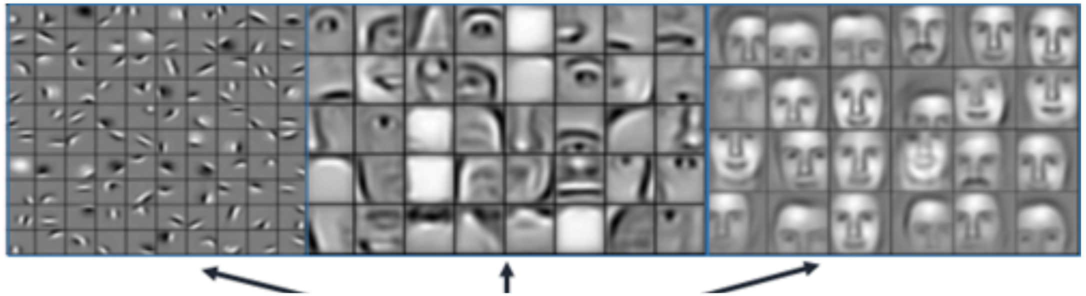
parameters of layer *L*

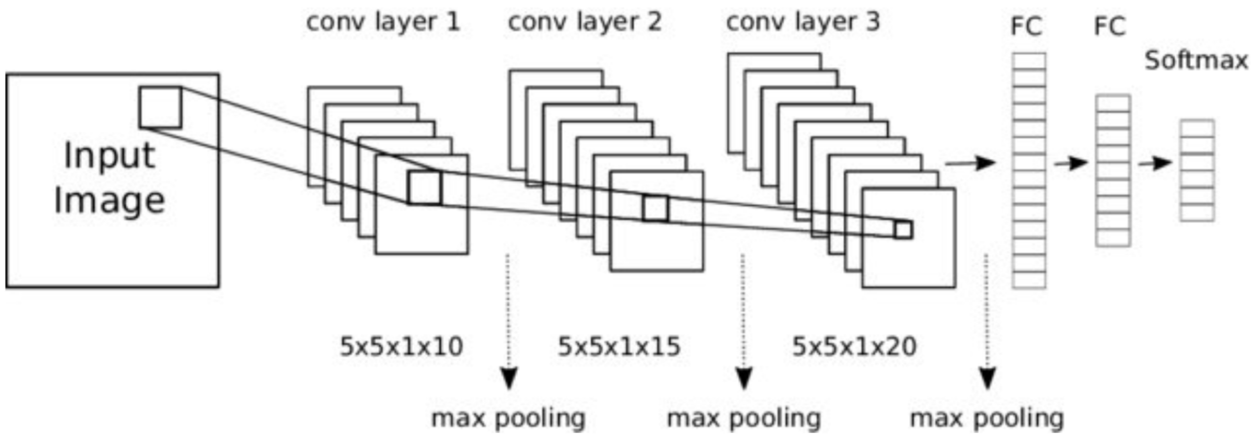
Deep Neural Networks

- Deep neural networks are generally better than other ML methods on images
- The series of layers between input and output compute relevant features automatically in a series of stages, just as our brains seem to.

Deep neural networks learn in multiple levels of representation, corresponding to different levels of abstraction

https://towardsdatascience.com

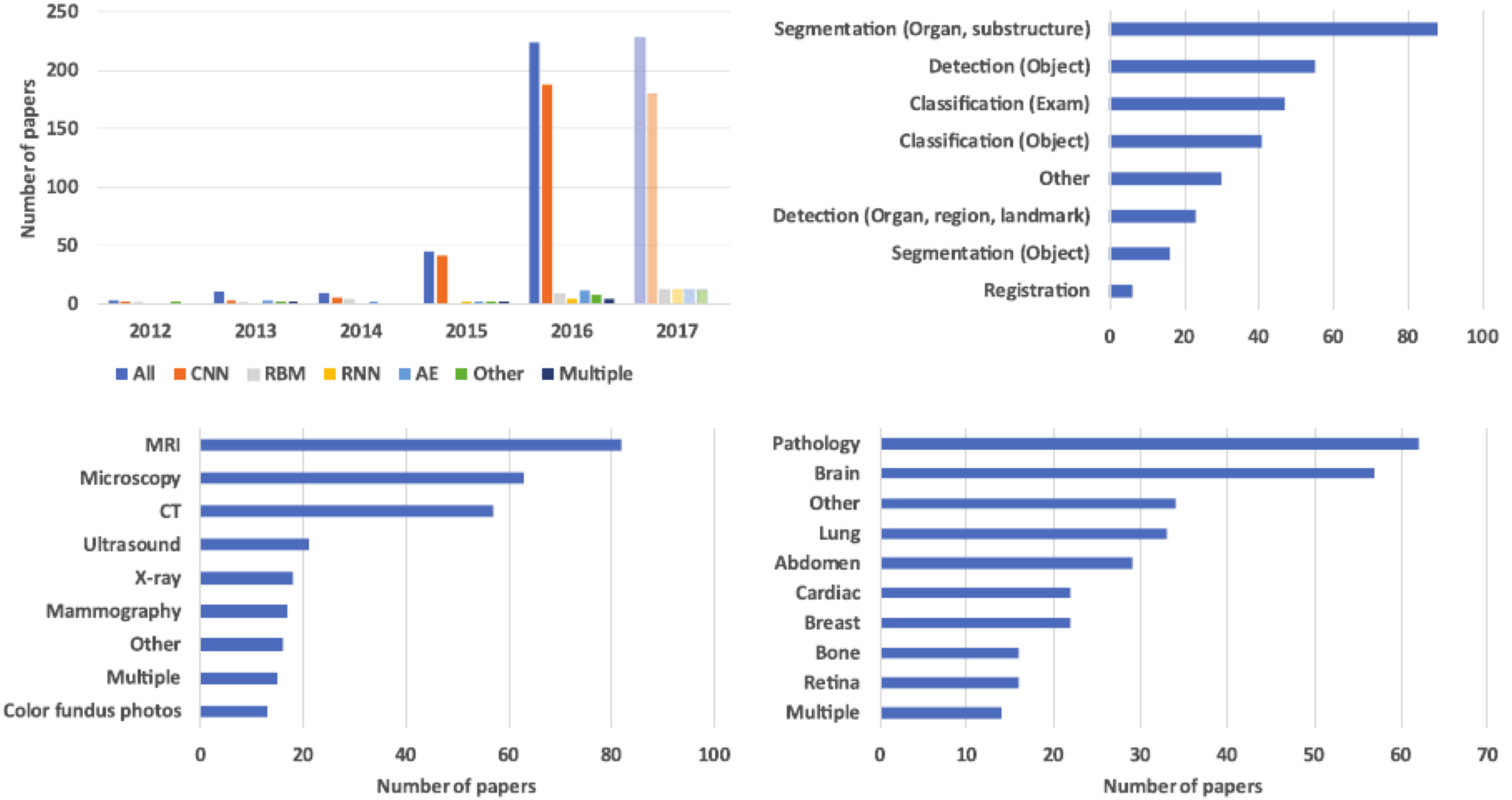




See demo code on https://github.com/retico/cmepda_medphys L10_code/Lecture10_demo_train_CNN.m

Deep Learning has become very popular in Medical Imaging

G. Litjens et al., A survey on in deep learning medical image analysis, Medical Image Analysis 42 (2017) 60–88



Breakdown of scientific papers by publication year, task addressed, imaging modality, and application area

DL vs. traditional ML approaches

- Deep Neural Networks can replace traditional handcrafted feature extraction

$\longrightarrow$ ***Data driven decision making***

- **Pros:**
  - No prior selection of problem-related feature => no loss of information
- **Cons:**
  - **Larger annotated** data samples are necessary
  - **Deep Neural Networks are black boxes: which image features are relevant for discrimination?**

Data augmentation (flip, rotate, scale images to augment data sets)

Model interpretability, explainable AI

References and sources

- References-Books

  - Bourlard, H.; Kamp, Y. (1988), Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics*. **59** (4–5): 291–294
  - Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1(4):541-551, Winter 1989
  - Teuvo Kohonen. Self-Organization of very large document collections: State of the art (1998)
  - S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice-Hall
  - I. Goodfellow, Y. Bengio, A. Courville. Deep Learning, The MIT Press
  - A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow, O'Reilly

- Sources

  - https://www.deeplearningitalia.com/una-panoramica-introduttiva-su-deep-learning-e-machine-learning/
  - http://www.asimovinstitute.org/neural-network-zoo/
  - https://towardsdatascience.com