UCK358E – INTR. TO ARTIFICIAL INTELLIGENCE

SPRING ‘23

LECTURE 1

INTRODUCTION

Instructor: Asst. Prof. Barış Başpınar

What are all these buzzwords?

Artificial Intelligence

Machine Learning

Big Data

Deep Learning

Neural Networks

Data Science

Pattern Recognition

Data Mining

Here is the dictionary!

Deep Learning = Neural Networks

Neural Networks ⊂ Machine Learning

Machine Learning ≈ Artificial Intelligence

All other buzzwords ≈ Machine Learning

Timeline of Ups and Downs

AI: Artificial Intelligence ML: Machine Learning

- **1950's**: AI/ML is the great future.
- **1970's**: AI became a bad word.
- **1980's**: Neural Networks are the great future.
- **1990's**: Neural Networks are not that good.
- **2010's**: Neural Networks are great after all.
- **Now**: AI/ML is the great future.

Why Learning?

Jeopardy's Watson is a one-task machine. Big task, but one task. (2011)

What the buzzwords have in common

- The same core premise:
  - *Machine Learning*
  - *Artificial Intelligence*
  - *Data Mining*
  - *Pattern Recognition*

“Automated detection of a **pattern** based on the **data** ”

Example: Credit Approval

Given the data of an applicant:

age	23 years
gender	male
annual salary	\$30,000
years in residence	1 year
years in job	1 year
current debt	\$15,000

should we extend credit?

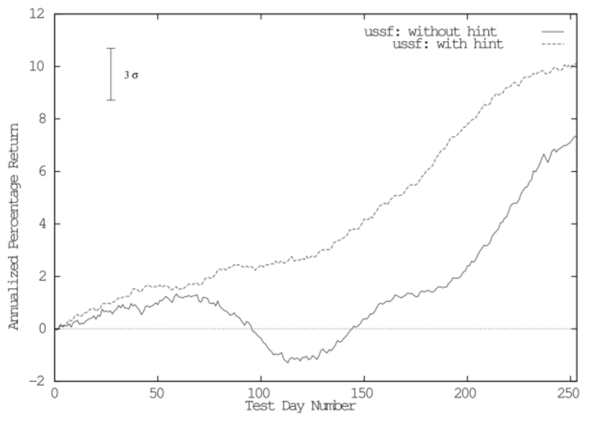
When should ML be used?

- ML is the technology of choice when:
  - A pattern exists.
  - We cannot pin it down mathematically.
  - We have a representative data set.

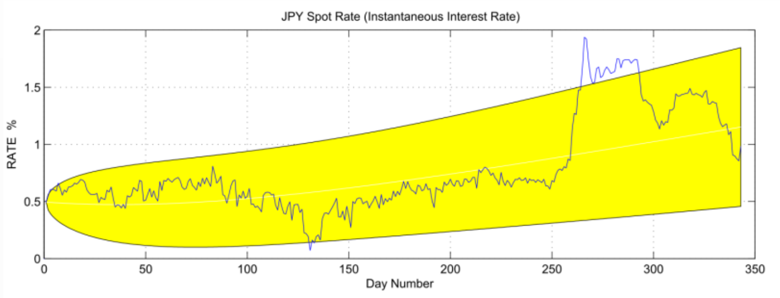
These criteria led to 3 waves of successful applications

1st Wave: Financial applications

- Market forecasting



- Financial model calibration



- Consumer and corporate credit assessment.

2nd Wave: E-commerce

- Recommender systems (Amazon, fashion, ...)

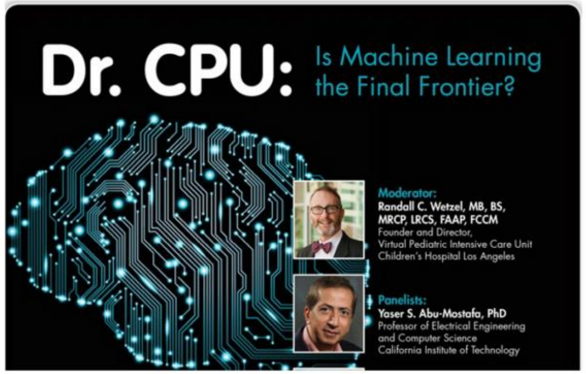
- Profiling

Famous ML e-commerce problem

US$1,000,000 Prize for the first 10% improvement (2006-2009)

3rd Wave: Medical Applications

- Medical diagnosis



- Data mining of medical records

The Growth of Machine Learning

- More Data:

  Enables us to pin down the pattern better.

  gradual → jump

- More Complex Models:

  Enables us to capture more complex patterns.

  gradual → jump

- More Computation:

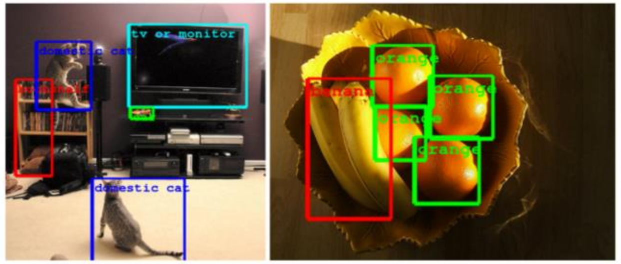
  Enables us to optimize very complex models.

  gradual → jump

ML Success Stories

The last years witnessed a huge surge in the practical impact of ML

speech recognition



object detection

machine translation

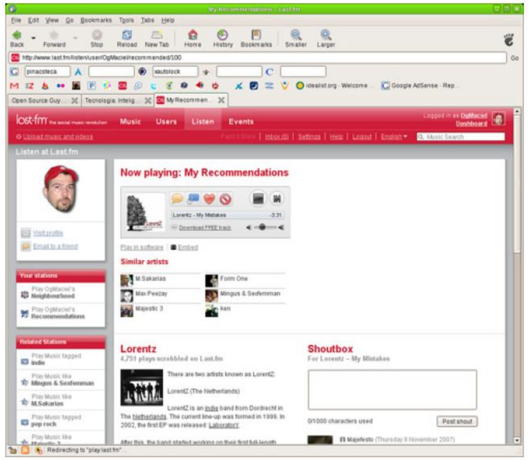
What happened?

A <u>qualitative</u> change in: data - models - computation

Different **data**: Total Profiling

Using multiple data sources:

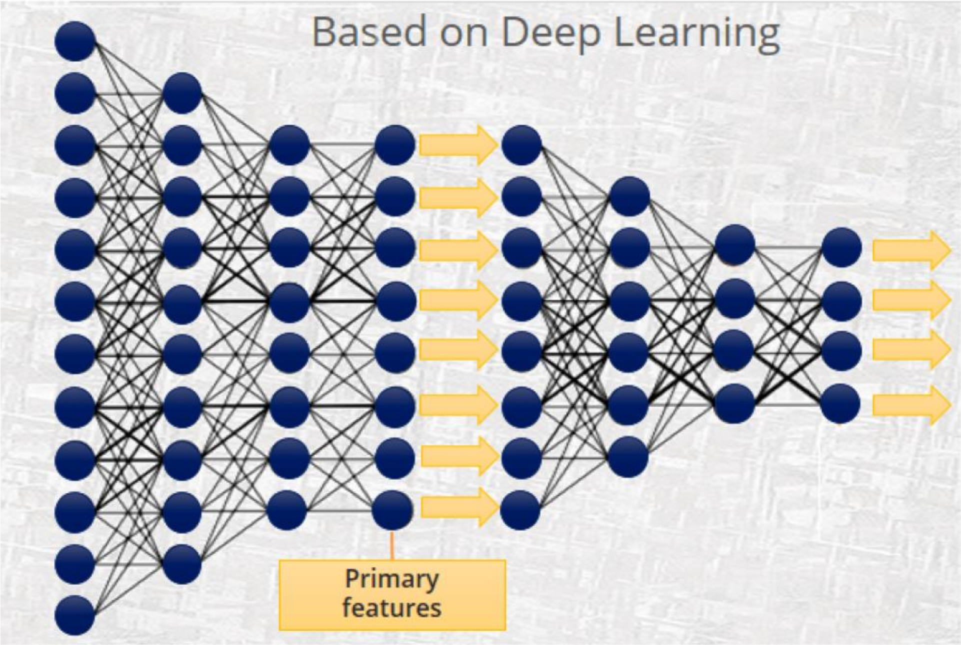
Movie preferences, Facebook posts, Amazon purchases, etc.



to profile a person.

What about the **models**?

A deep Neural Network with millions of parameters



Automated Feature Extraction

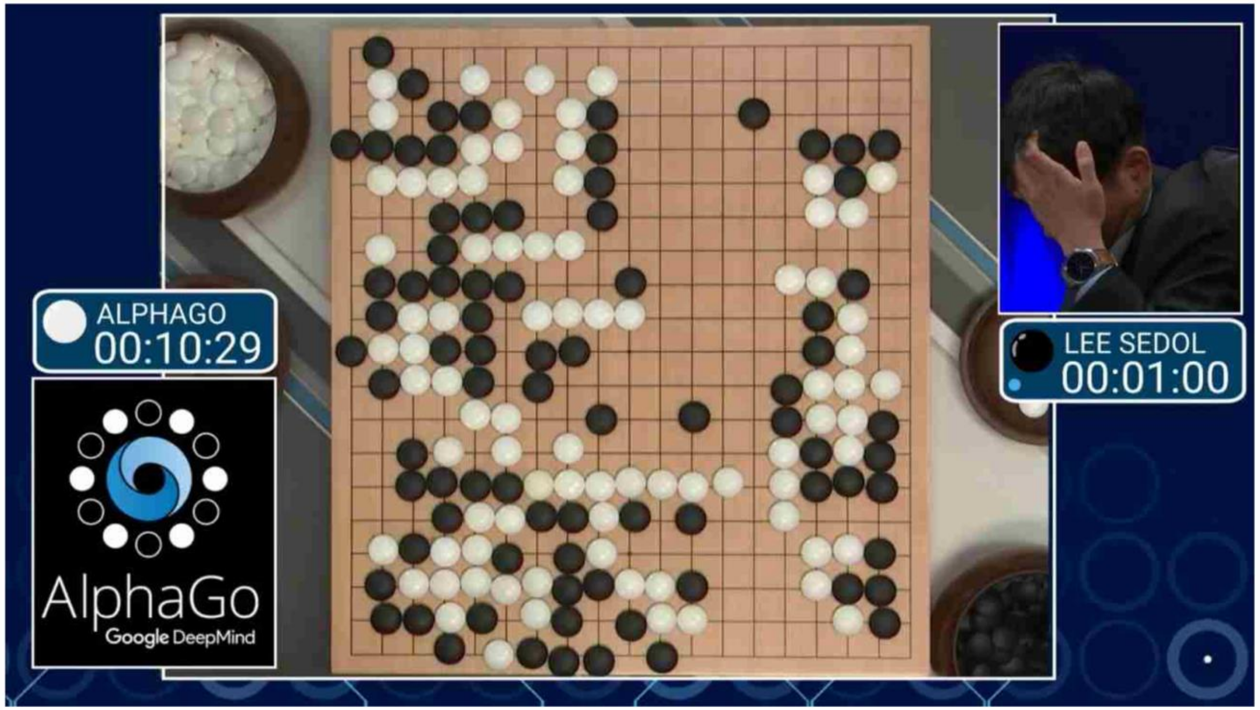
Jump in **computation** speed

Commercially available specialized hardware

Gain in ML speed is more than 2 orders of magnitude

Real “Intelligence” achieved

From: Replicating human skills   To: Beating human intelligence



ML system discovered <u>novel</u> strategic moves in the game of **Go** (2015)

Machine Learning Perspective

Data Collection

- Measurement Devices
- Sensor
- Images
- DBs

Preprocessing

- ***Feature Extraction***
- Noise Filtering
- Normalization

Dimensionality Reduction (Optional)

- ***Feature Selection***
- Feature Projection

Model Learning

- Classification
- Regression
- Clustering
- Description

Model Testing

- Cross Validation
- Bootstrap

Learning Essentials

- ***Machine Learning:*** Data -> Train a Model on Data -> Make Predictions on New Data

- ***Feature Engineering:*** Extracting useful patterns from data that will be used for ML models for classification

- ***Feature Learning:*** Feature learning algorithms find the common patterns that are important to classify samples and extract them automatically to be used in a classification or regression process.

- ***Deep Learning:*** New methods and strategies designed to generate deep hierarchies of non-linear features so that Deep architectures with dozens of layers of non-linear hierarchical features can be trained

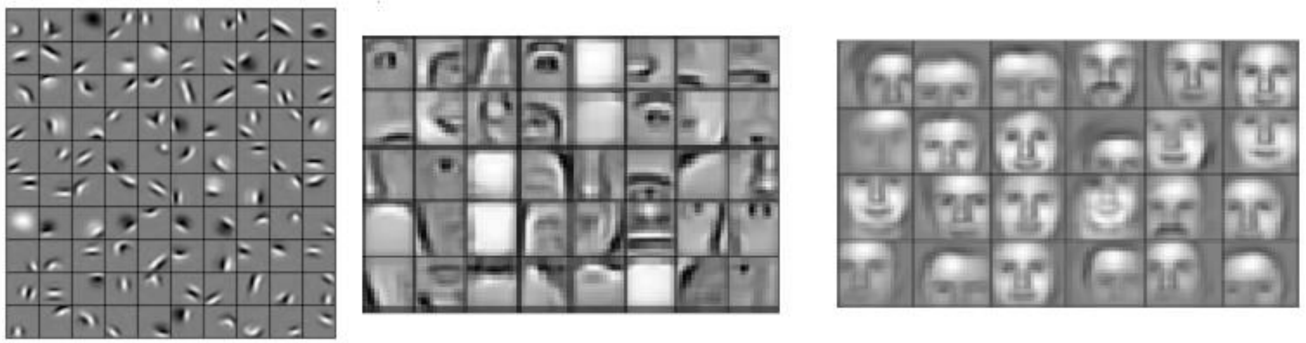
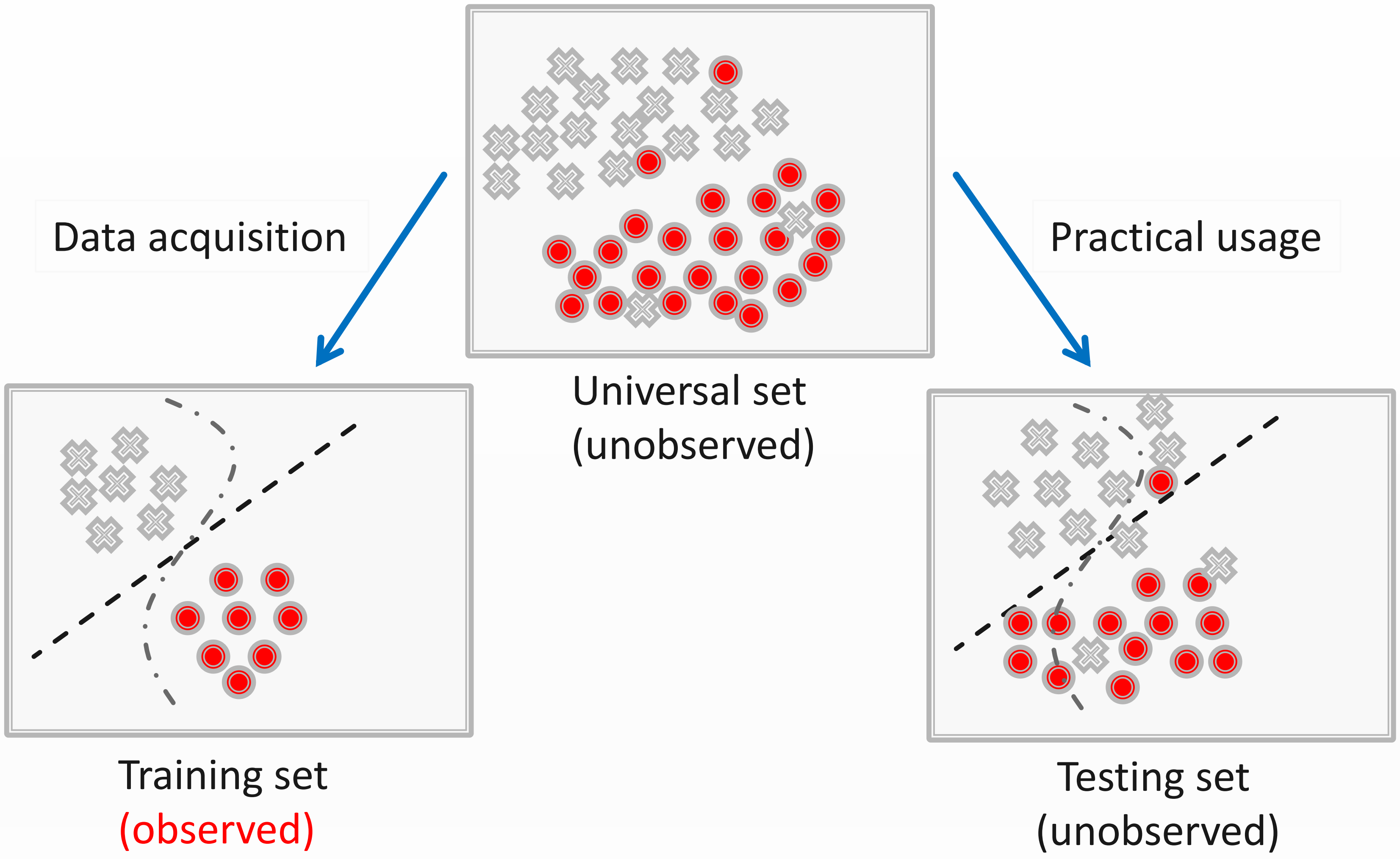


Image by Honglak Lee and colleagues (2011) , “Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks”

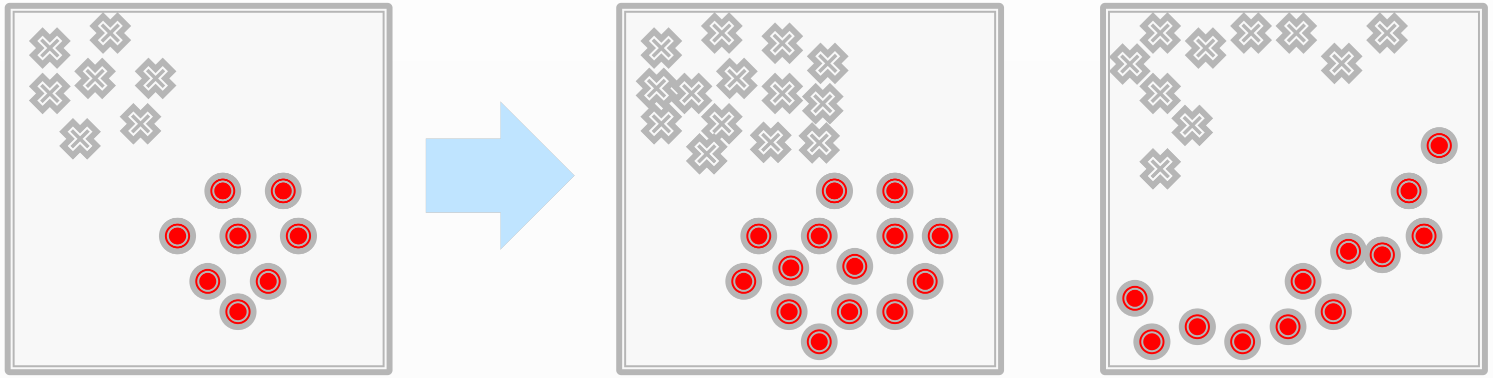
Training and testing



Training and testing

- Training is the process of making the system able to learn.

- No free lunch rule:
  - Training set and testing set come from the same distribution
  - Need to make some assumptions or bias



Performance

- There are several factors affecting the performance:
  - **Types of training** provided
  - The form and extent of any initial **background knowledge**
  - The **type of feedback** provided
  - The **learning algorithms** used

- Two important factors:
  - Modeling
  - Optimization

Algorithms

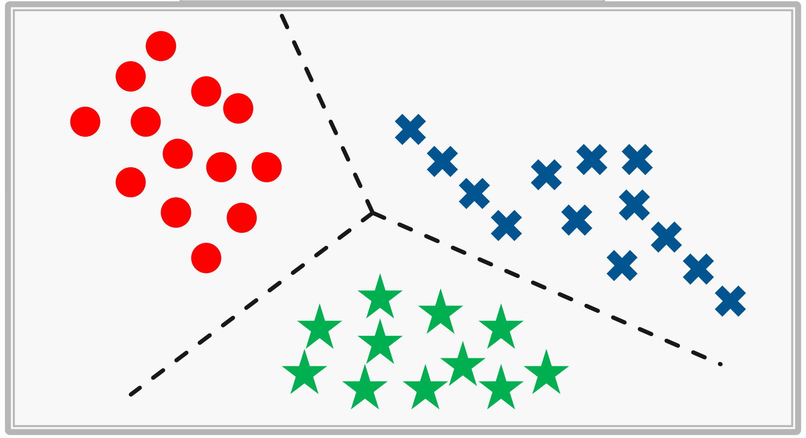
- The success of machine learning system also depends on the algorithms.
- The algorithms control the search to find and build the knowledge structures.
- The learning algorithms should extract useful information from training examples.

Algorithms

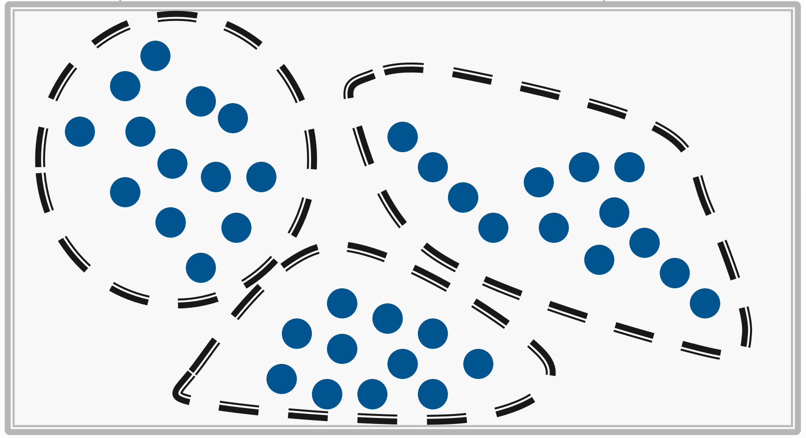
- **Supervised learning** ($\{x_n \in R^d, y_n \in R\}_{n=1}^N$)
  - Prediction
  - Classification (discrete labels), Regression (real values)
- **Unsupervised learning** ($\{x_n \in R^d\}_{n=1}^N$)
  - Clustering
  - Probability distribution estimation
  - Finding association (in features)
  - Dimension reduction
- **Semi-supervised learning**
- **Reinforcement learning**
  - Decision making (robot, chess machine)

Algorithms

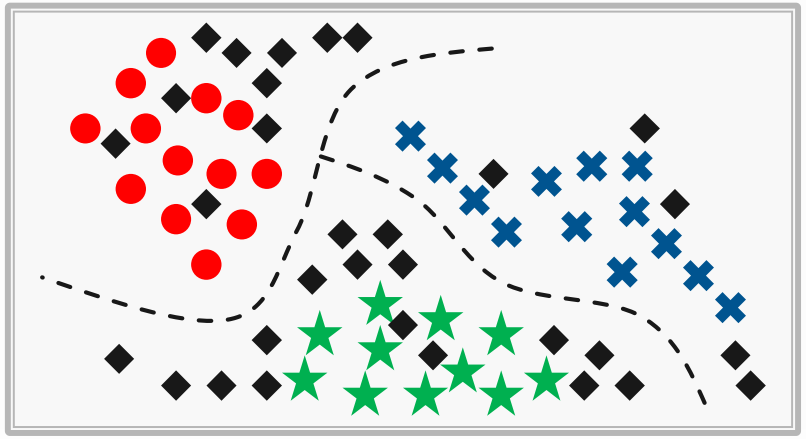
Supervised learning



Unsupervised learning

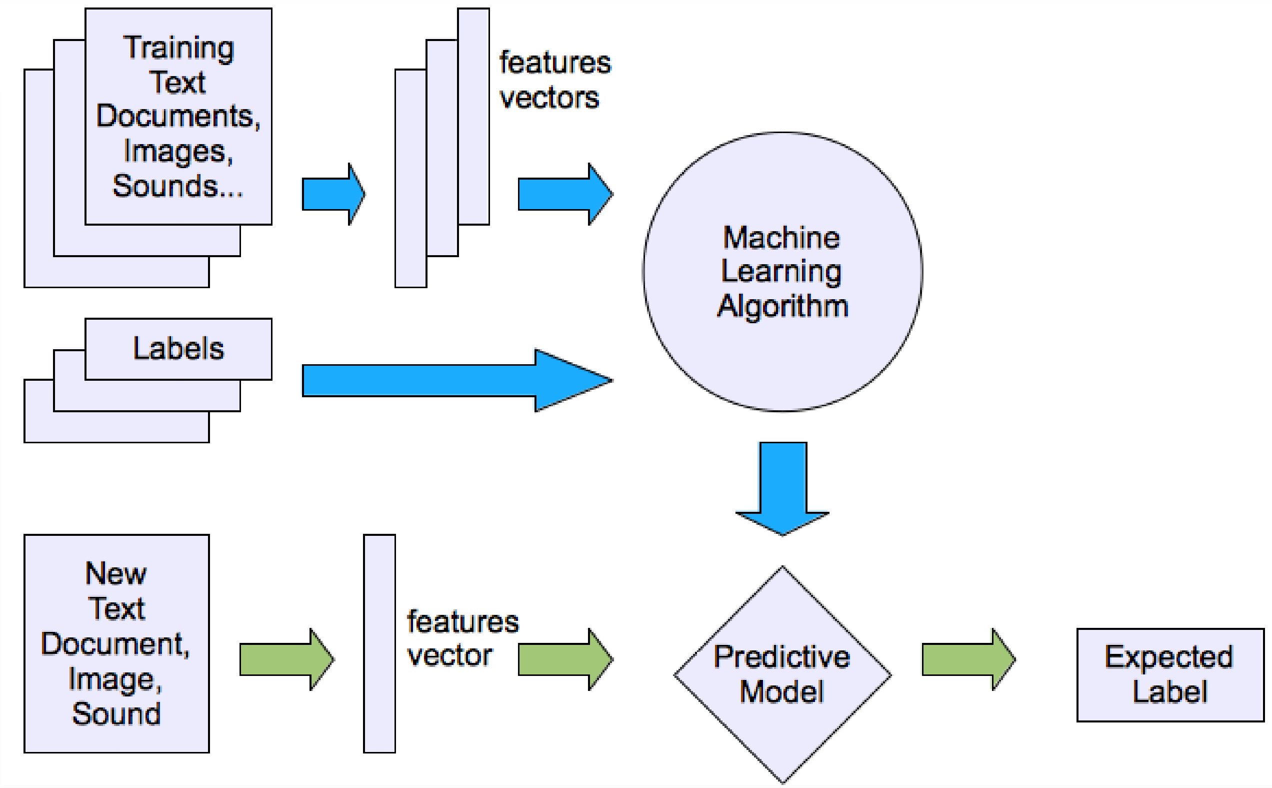


Semi-supervised learning



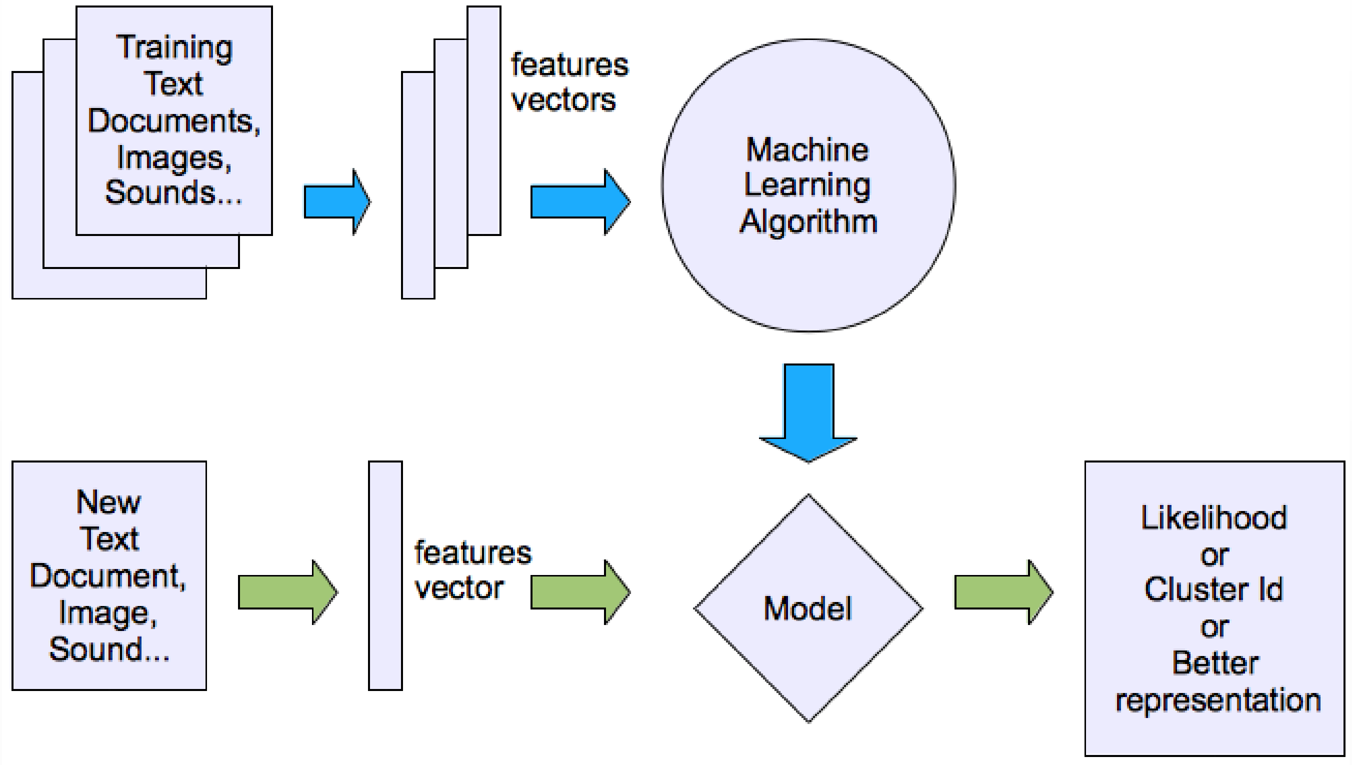
Machine learning structure

- Supervised learning



Machine learning structure

- Unsupervised learning



What are we seeking?

- Supervised: Low E-out or maximize probabilistic terms

$$error = \frac{1}{N} \sum_{n=1}^{N} [y_n \neq g(x_n)]$$

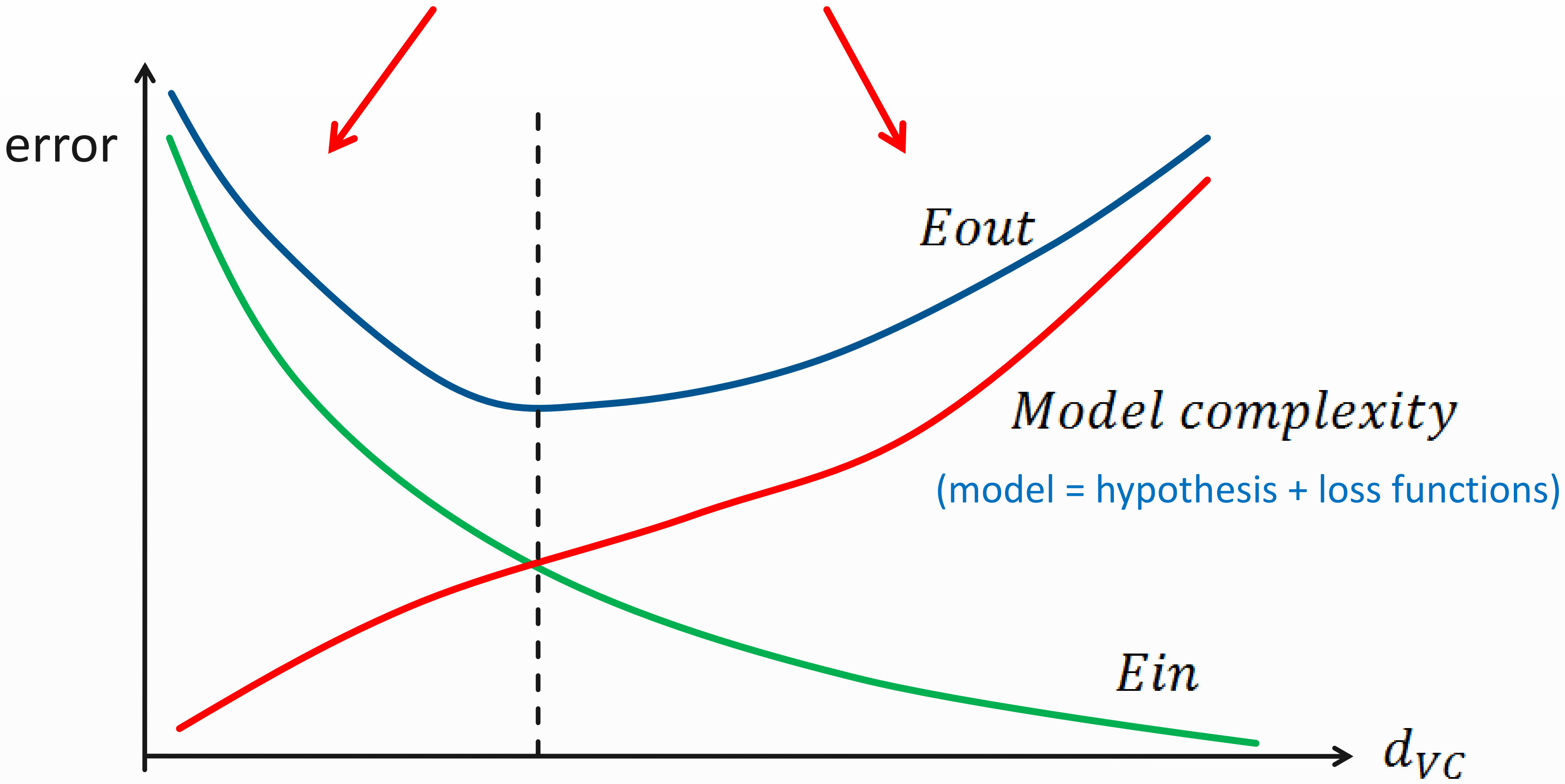
E-in: for training set
E-out: for testing set

$$Eout(g) \leq Ein(g) \pm O\left(\sqrt{\frac{d_{VC}}{N} \ln N}\right)$$

- Unsupervised: Minimum quantization error, Minimum distance, MAP, MLE(maximum likelihood estimation)

What are we seeking?

Under-fitting VS. Over-fitting (fixed *N*)

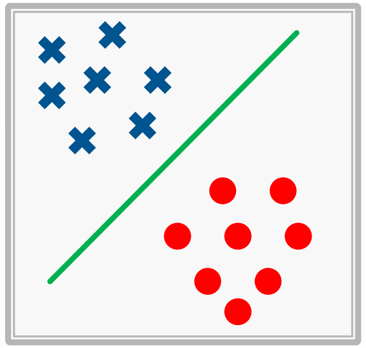


Learning techniques

- Supervised learning categories and techniques
  - **Linear classifier** (numerical functions)
  - **Parametric** (Probabilistic functions)
    - Naïve Bayes, Gaussian discriminant analysis (GDA), Hidden Markov models (HMM), Probabilistic graphical models
  - **Non-parametric** (Instance-based functions)
    - *K*-nearest neighbors, Kernel regression, Kernel density estimation, Local regression
  - **Non-metric** (Symbolic functions)
    - Classification and regression tree (CART), decision tree
  - **Aggregation**
    - Bagging (bootstrap + aggregation), Adaboost, Random forest

Learning techniques

- Linear classifier



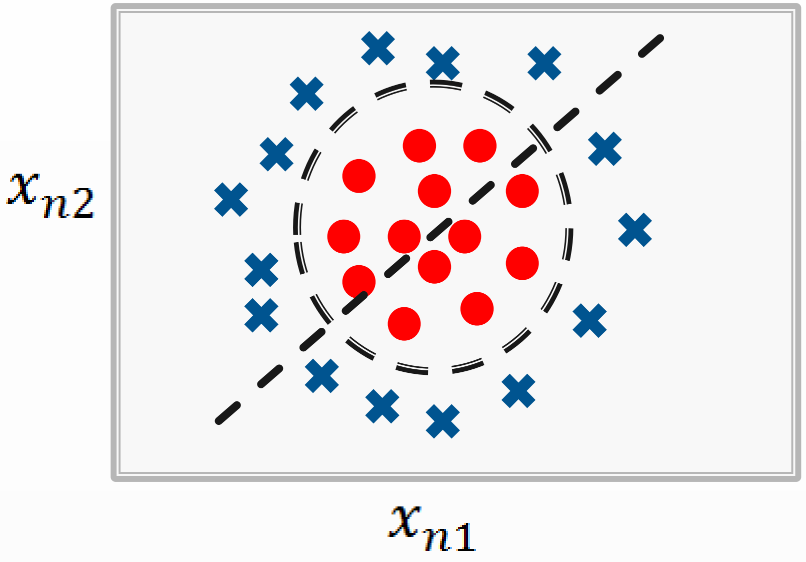
$$g(x_n) = sign(w^T x_n)$$

, where *w* is an *d*-dim vector (learned)

- Techniques:
  - Logistic regression
  - Support vector machine (SVM)
  - Multi-layer perceptron (MLP)

Learning techniques

- Support vector machine (SVM):



$$x_n = [x_{n1}, x_{n2}]$$

$$x_n = [x_{n1}, x_{n2}, x_{n1} * x_{n2}, {x_{n1}}^2, {x_{n2}}^2]$$

$$g(x_n) = sign(w^T x_n)$$

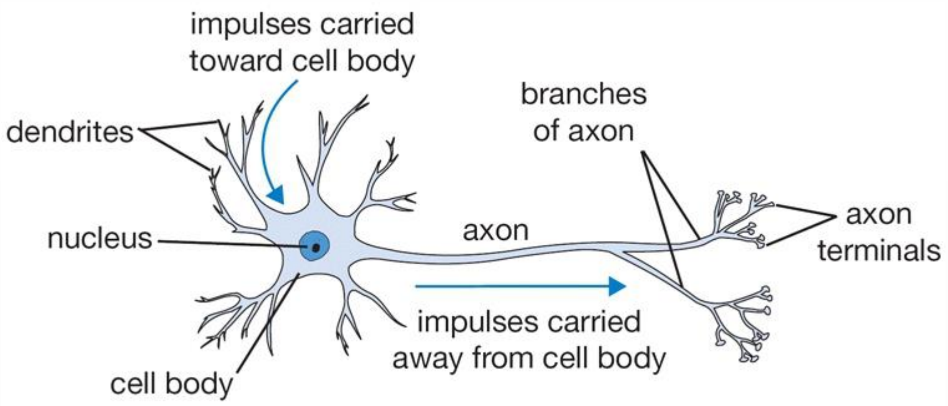
  - Non-linear case

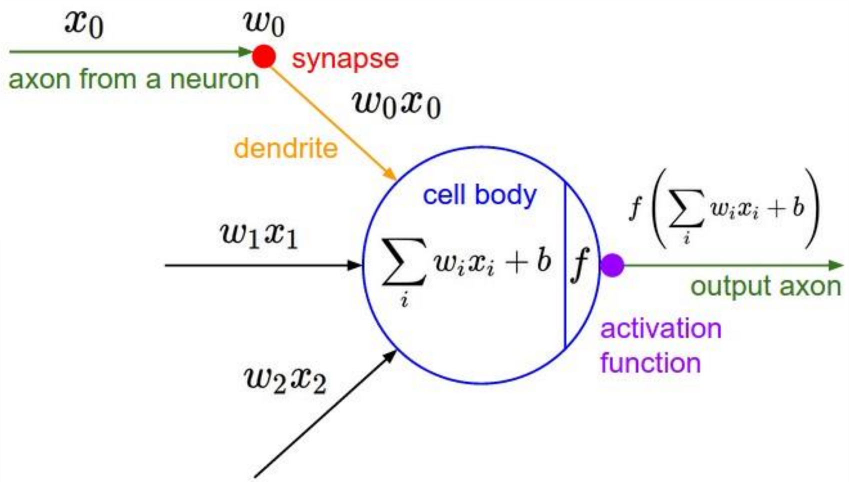
- Linear to nonlinear: **Feature transform** and **kernel function**

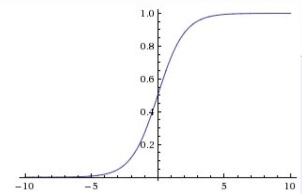
Learning techniques

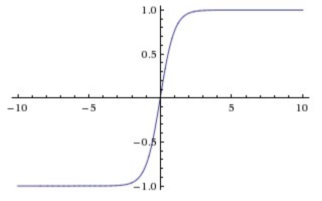
- Unsupervised learning categories and techniques
  - **Clustering**
    - K-means clustering
    - Spectral clustering
  - **Density Estimation**
    - Gaussian mixture model (GMM)
    - Graphical models
  - **Dimensionality reduction**
    - Principal component analysis (PCA)
    - Factor analysis

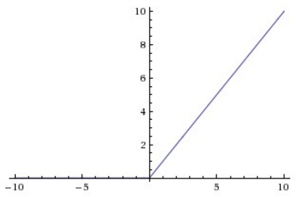
Deep Learning Essentials











```python
class Neuron(object):
  # ...
  def forward(inputs):
    """ assume inputs and weights are 1-D numpy arrays and bias is a number """
    cell_body_sum = np.sum(inputs * self.weights) + self.bias
    firing_rate = 1.0 / (1.0 + math.exp(-cell_body_sum)) # sigmoid activation function
    return firing_rate
```

from: CS231n Convolutional Neural Networks for Visual Recognition, http://cs231n.github.io/neural-networks-1/

Neural Network Architectures

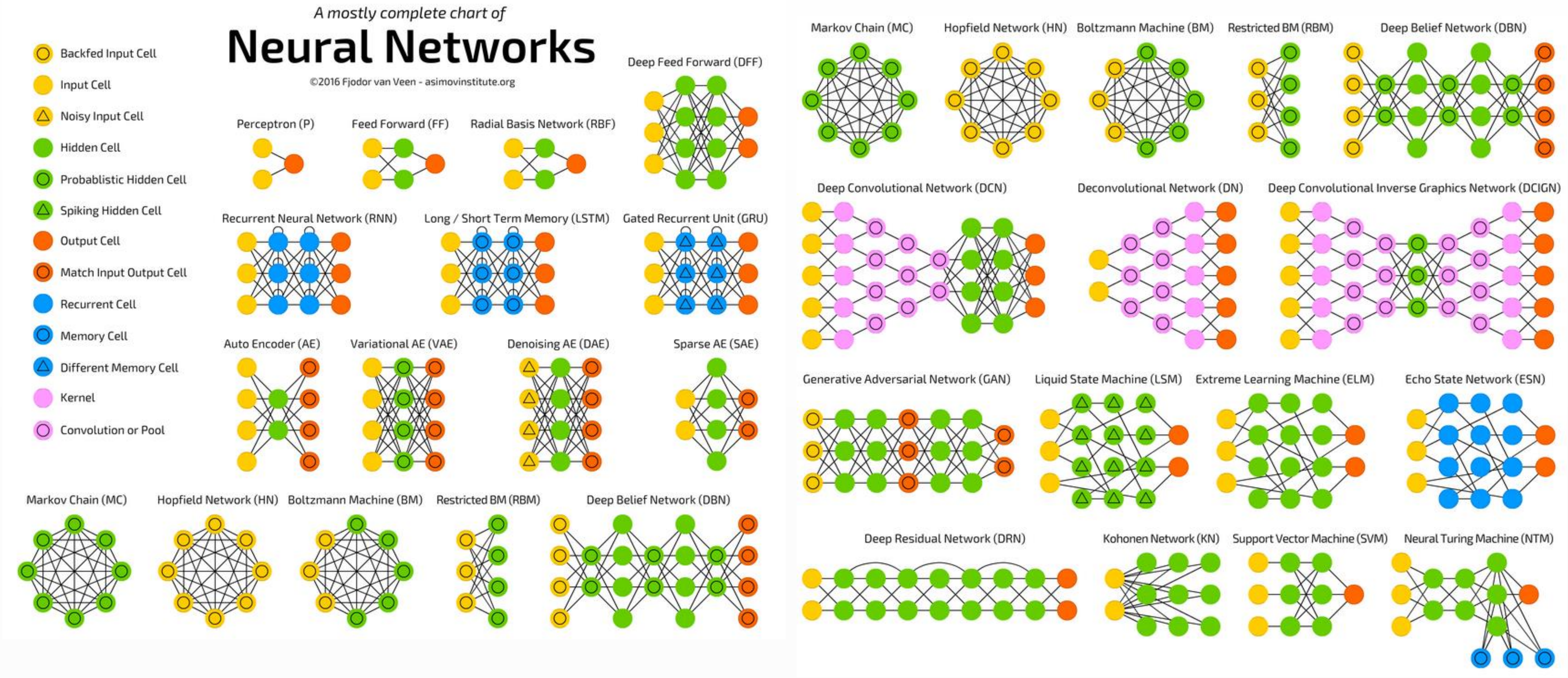


Image: The Asimov Institute – The Neural Network Zoo - http://www.asimovinstitute.org/neural-network-zoo/

Applications

- Face detection
- Object detection and recognition
- Image segmentation
- Multimedia event detection
- Economical and commercial usage

References

- W. L. Chao, J. J. Ding, “Integrated Machine Learning Algorithms for Human Age Estimation”, NTU, 2011.
- Y. S. Abu-Mostafa, “Artificial Intelligence: Evolution & Revolution (& Hype!)”, 2018.