Introduction to Machine Learning (Neural Networks)

Schedule

- I. Introduction
- II. Logistic regression (review)
- III. Neural network
- VI. Convolutional NN
- V. Other

Related Terms

Deep Learning

Neural Networks

Artificial Intelligence

Data Science

Machine Learning

Data Mining

Pattern Recognition

Statistical Modeling

Related Terms

Game Theory

STT

MT

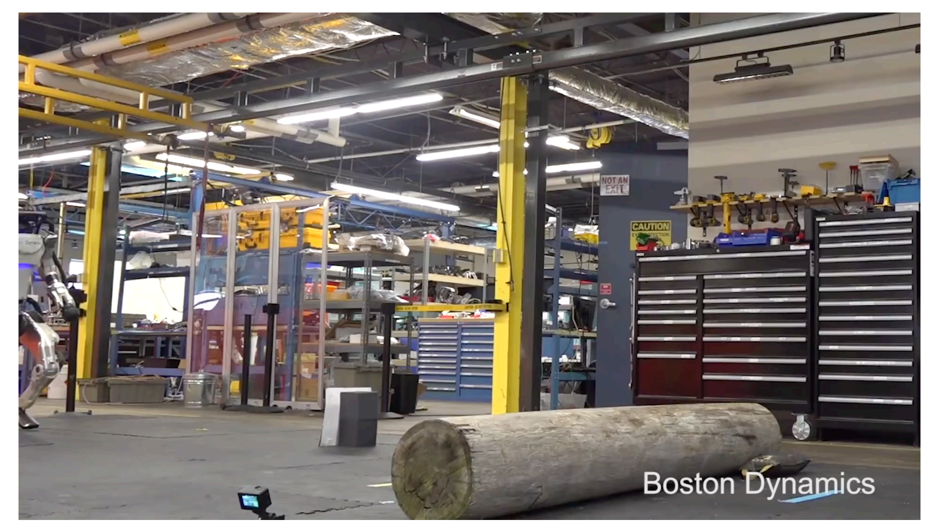
Natural Language Processing

Artificial Intelligence

Robotics

Planning

Machine Learning



Logic Reasoning

Related Terms

Algorithmic Trading

NoSQL

KDD

Data Science

Machine Learning

KDD

Database

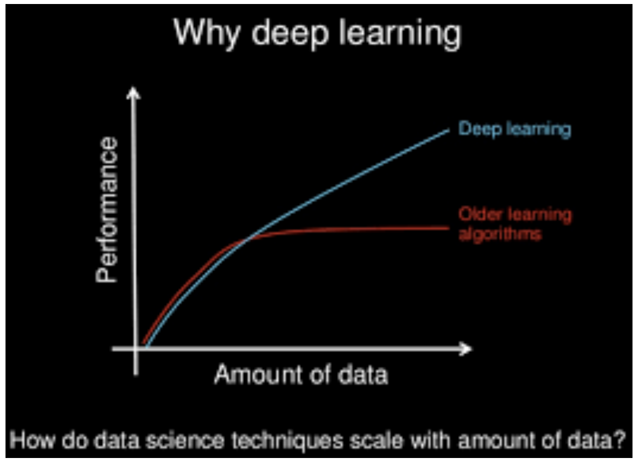
Data Mining

Recommendation System

Business Intelligence

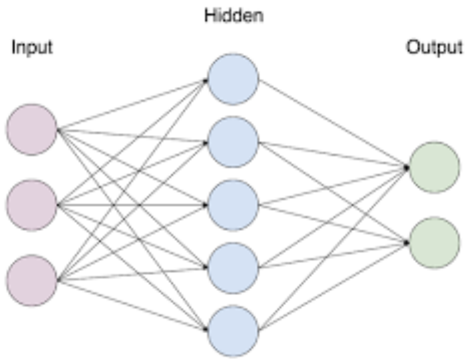
Data Management

Related Terms



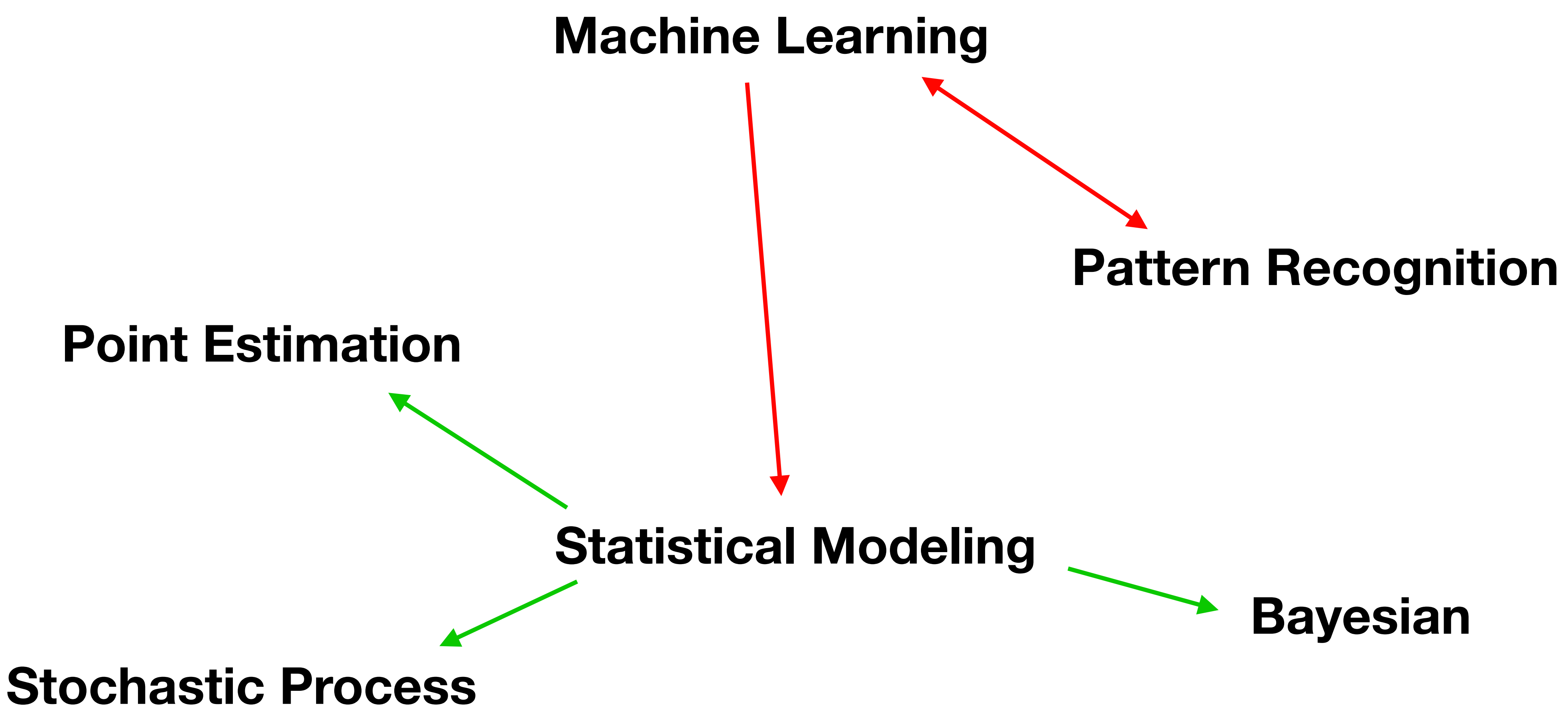
Deep Learning

Neural Networks



Machine Learning

Related Terms



Related Terms

Deep Learning

Neural Networks

Artificial Intelligence

Data Science

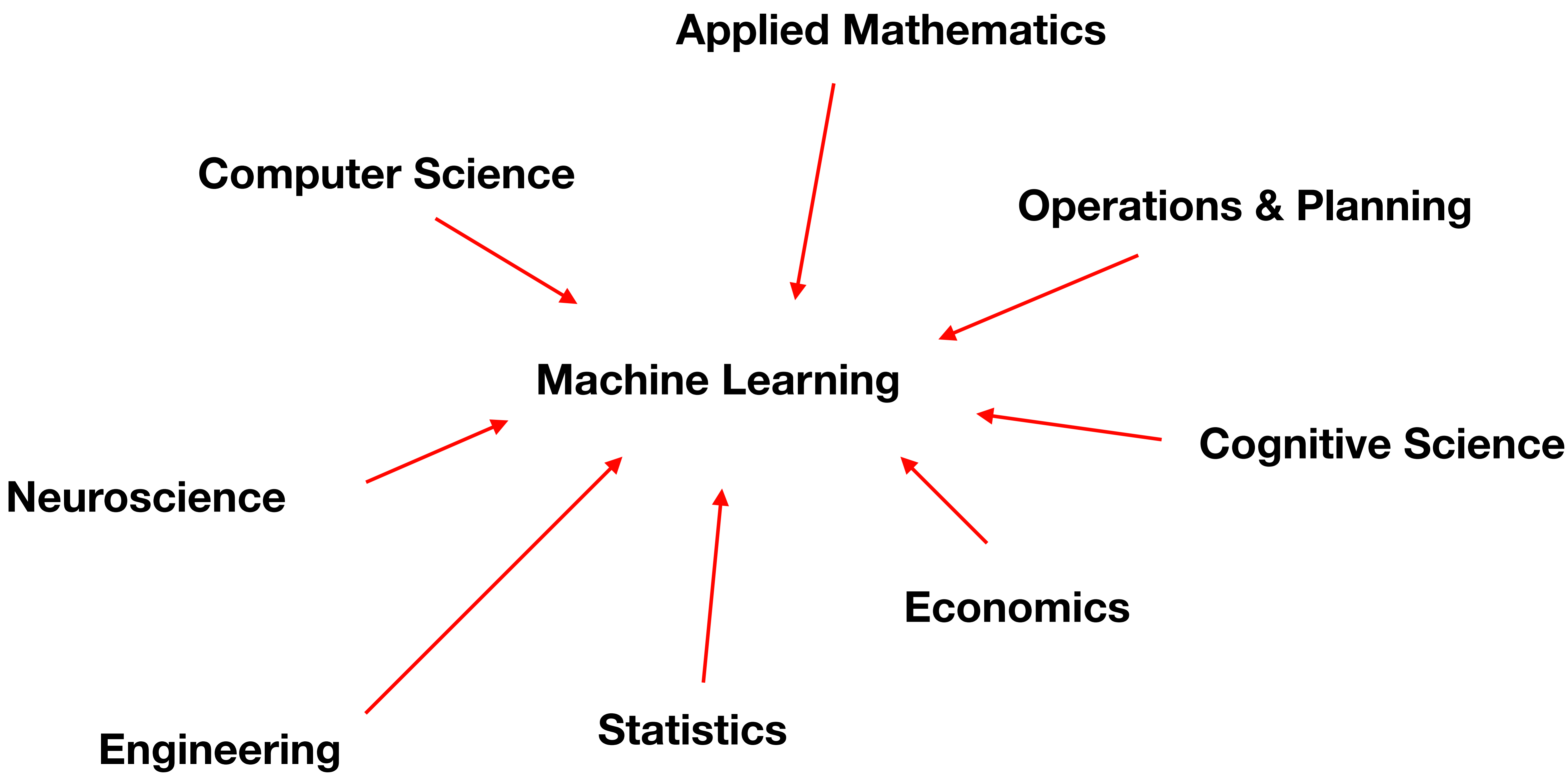
Machine Learning

Data Mining

Pattern Recognition

Statistical Modeling

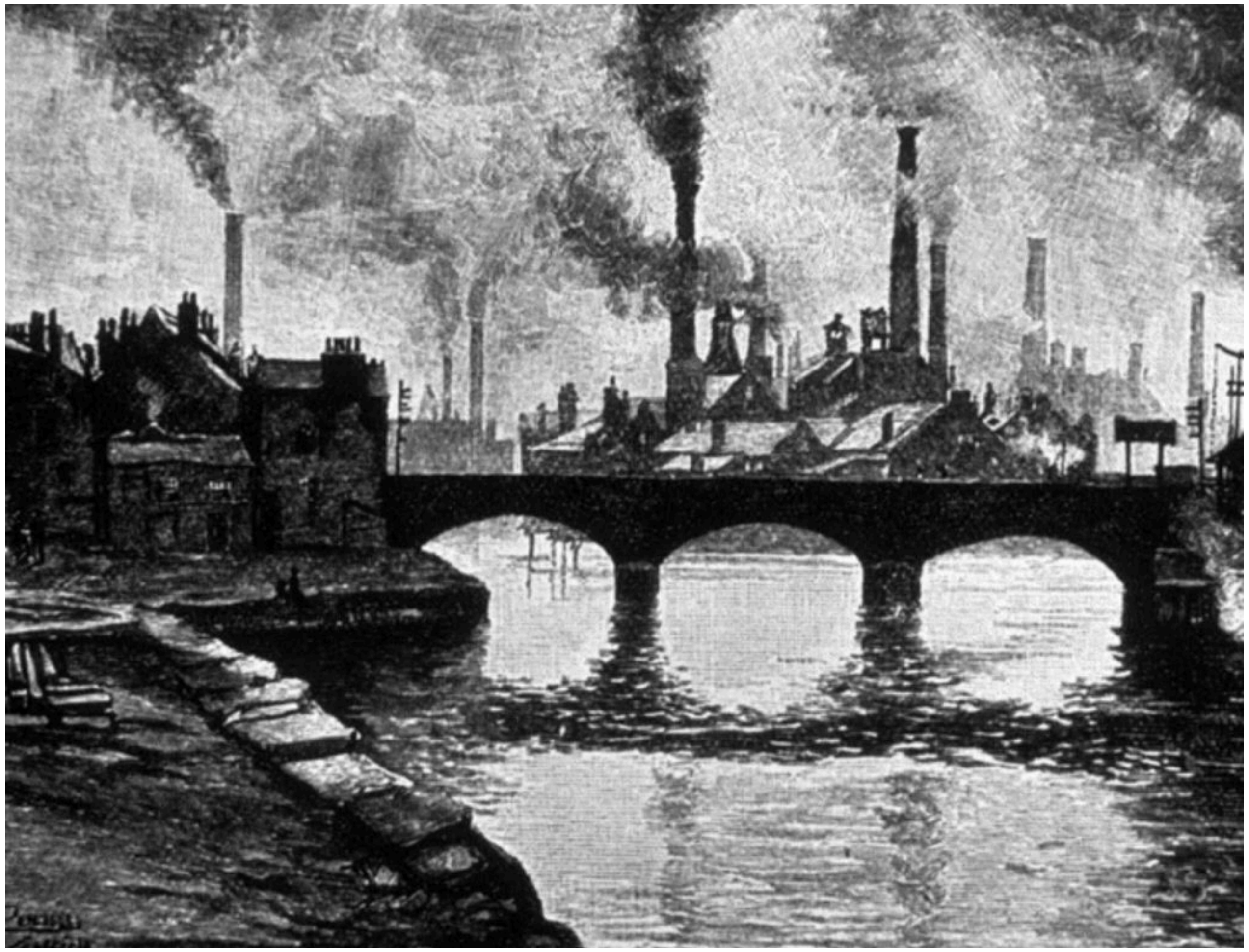
RELATED Fields



Information Revolution Era

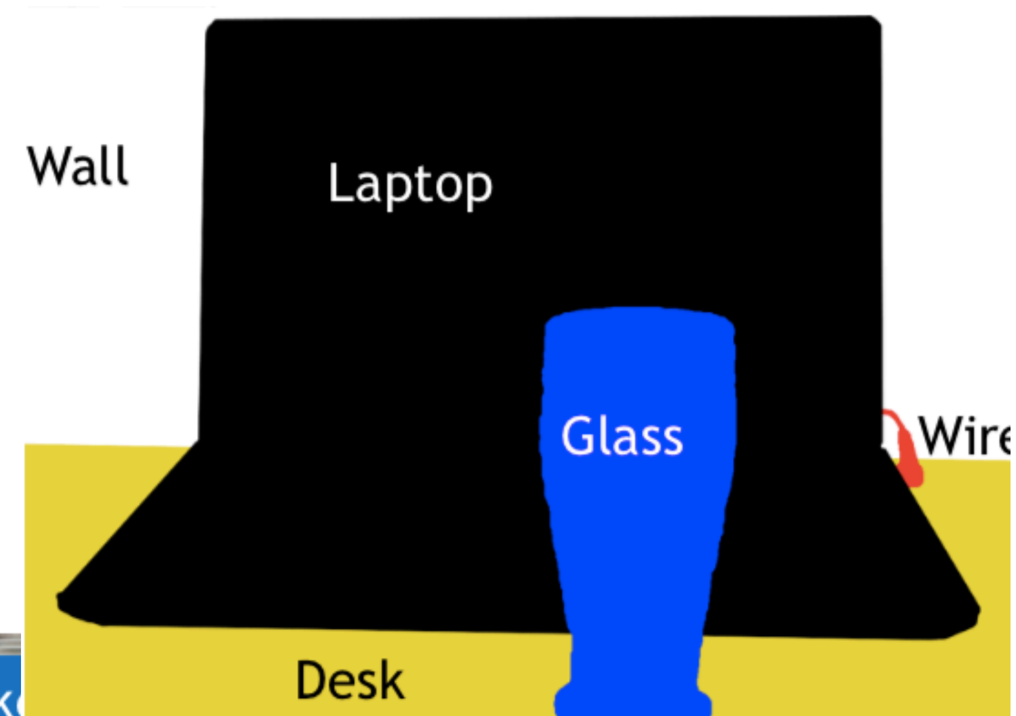
We have abundant of data:

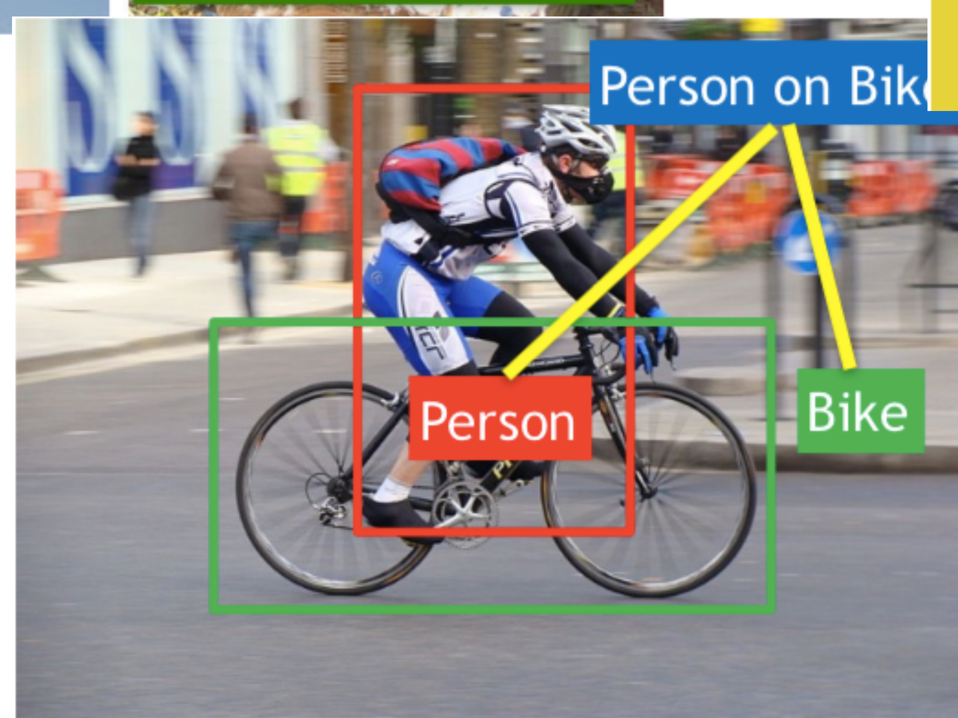
- **People: the web, blog, social network, government**
- **Science: biomedical data, climate data, scientific literature**
- **Business: e-commerce, e-trading, advertising, personalisation.**



We need tools for understanding, and decision-making from the data.

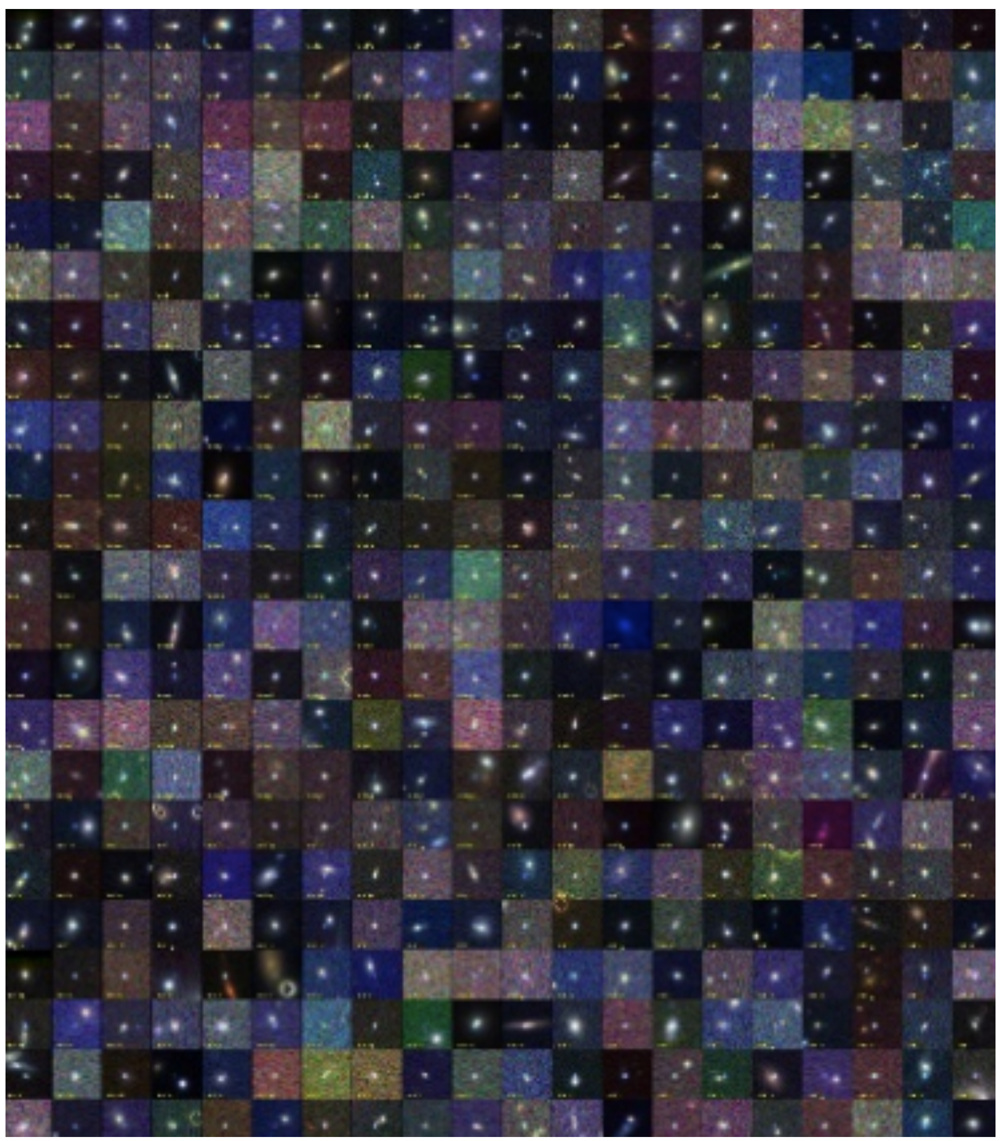
Application: Computer Vision

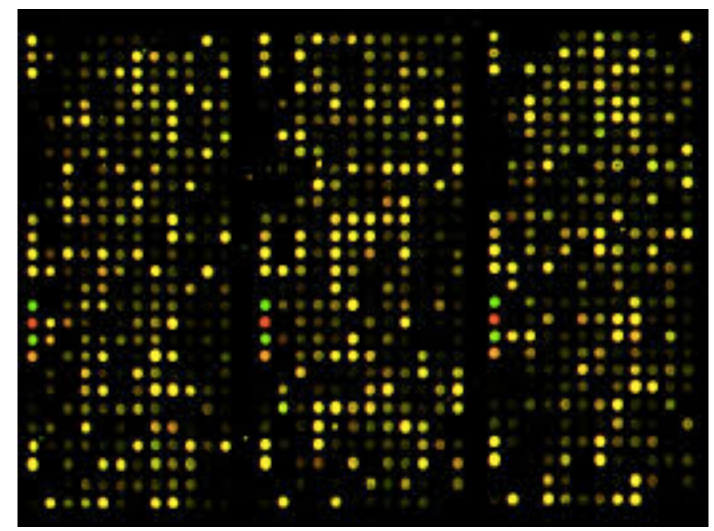




- **Detection(face/object)**
- **Recognition**
- **Captioning**
- **Segmentation**
- **...**

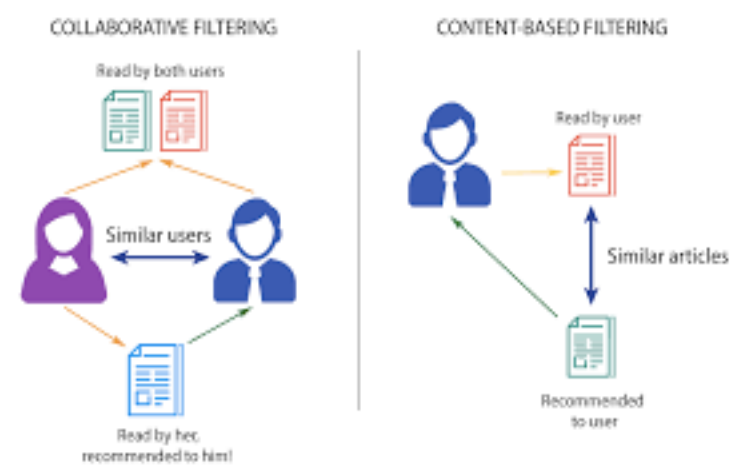
Application: biomedical & astronomy





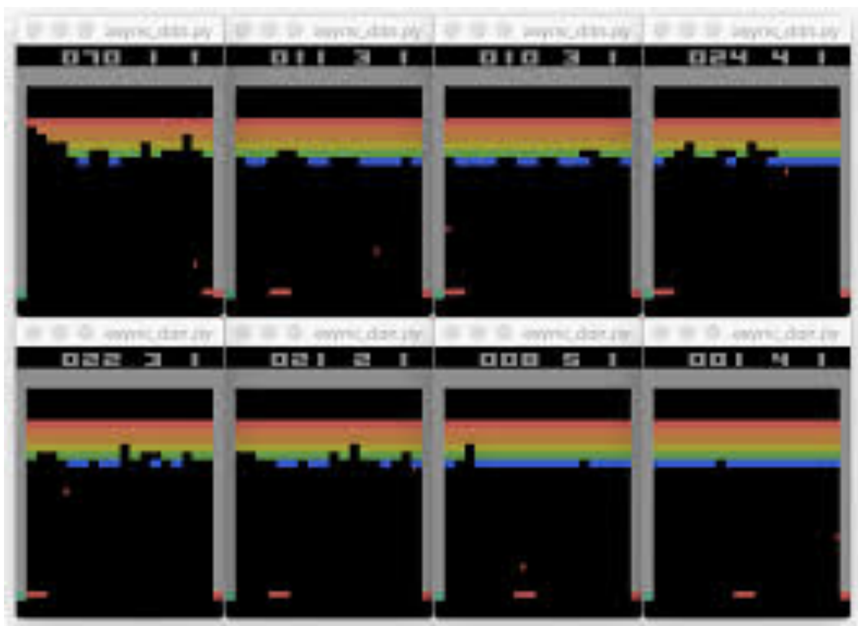
- **Microarray(compressed sensoring)**
- **astronomy**

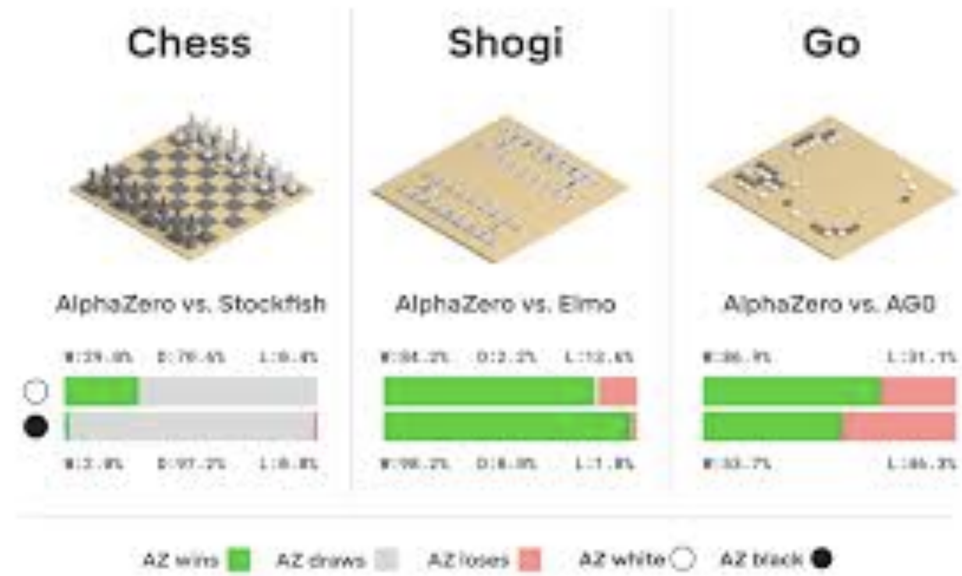
Application: Recommendation System

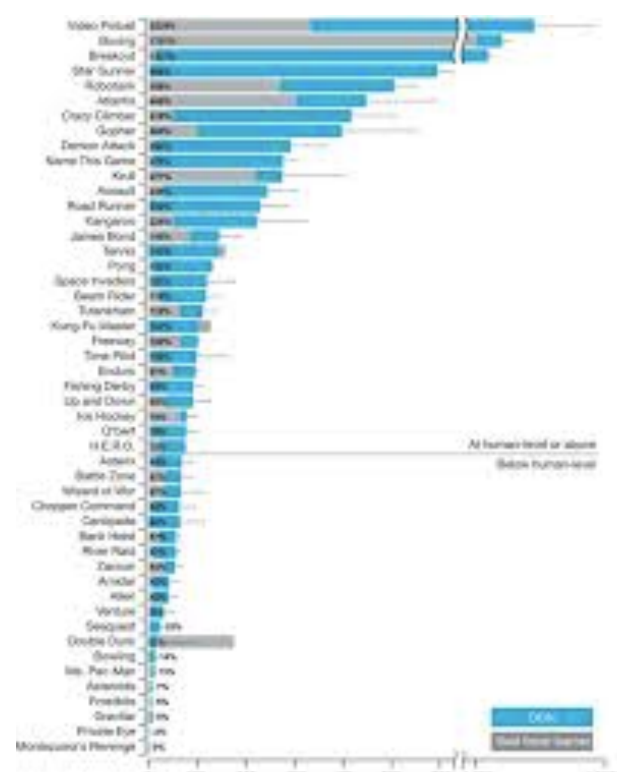


- **Recommendation system is everywhere.**
- **It is one of of the backbone of Internet.**

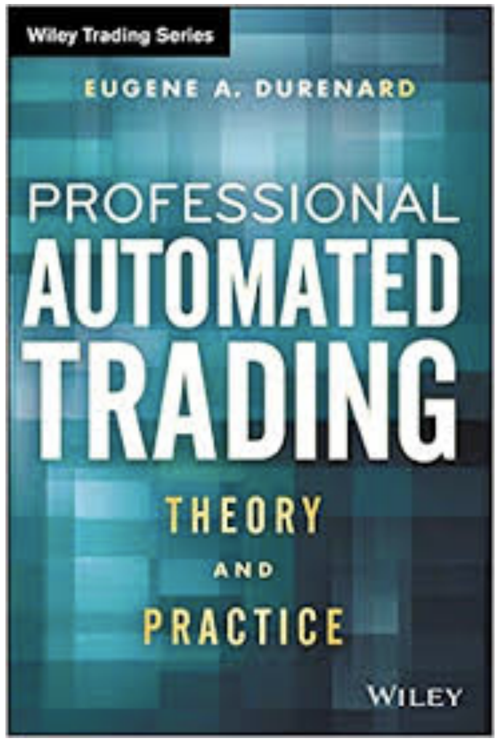
Application: Games







Application: Automated Trading



Types of Learning

- The machine is given a series of input:

$$x_1, x_2, x_3, \ldots$$

- Supervised learning: the machine is also given a corresponding output $y_1, y_2, y_3, \ldots$. The task is to learn the correct output given the input.
- Unsupervised learning: the machine is ask to build a model for prediction, decision making, etc.
- Reinforcement learning: the machine can perform an action $a_1, a_2, a_3, \ldots$ based on the input to update the world, the world can give back some rewards $r_1, r_2, r_3, \ldots$. The task here is usually to maximize the rewards.

Typical Problems

There are many applications of machine learning, but we often see application can be modeled around a few problems. Here are some:

- Classification
- Regression
- Clustering
- Dimensionality Reduction

Classification

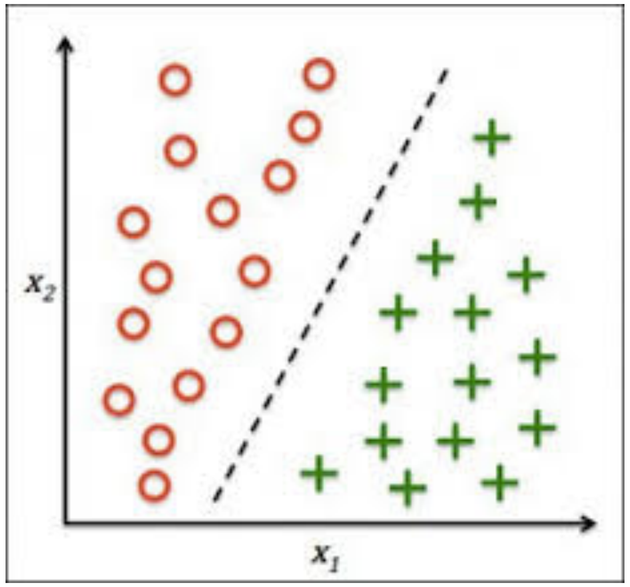
- The data is represented in some vector space.
- Let's say we have input data.

$$x = (x_1, x_2, x_3, \ldots)$$

- For each of the input data, it's represented by a vector. We call it features.
- For each of the input data, we know their discrete class labels.

$$y = (y_1, y_2, y_3, \ldots)$$

- The goal is to give the correct label for some given new data.

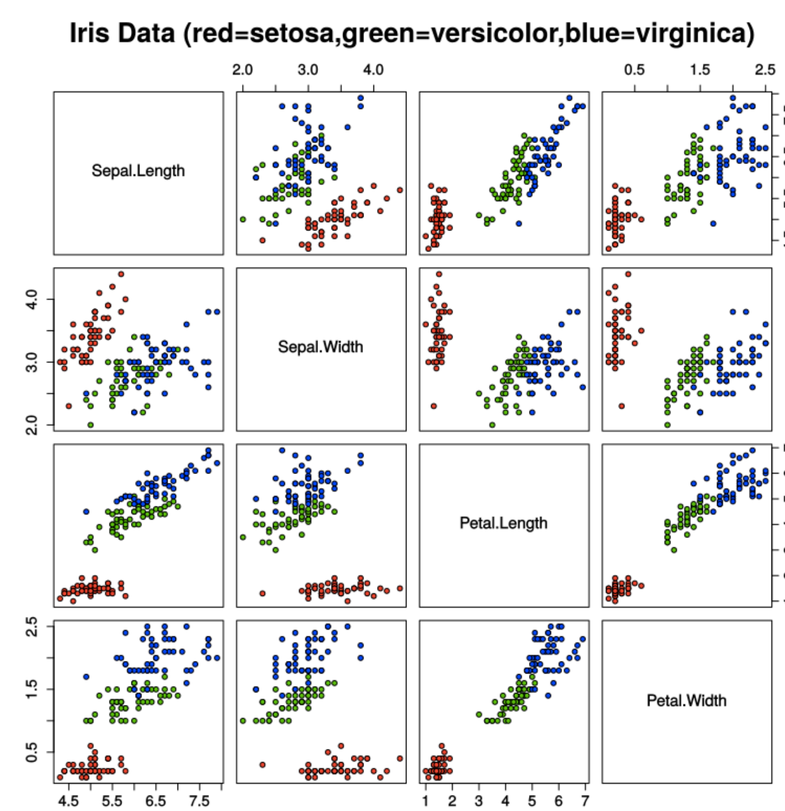


Example:
- Healthy or ill?
- Spam or normal email?
- Is it an apple, a pear, or an orange?

Classification Example: Iris Data

- 4 features, 3 classes, and 150 instances.
- The features are the sepal length, sepal width, petal length, and petal width in cm.
- The classes are three iris species: setosa, versicolor, and virginica.
- The data is collected by Anderson (1935) and used by Fisher (1936) with linear discriminant function.

```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
...
7.0,3.2,4.7,1.4,Iris-versicolor
6.4,3.2,4.5,1.5,Iris-versicolor
6.9,3.1,4.9,1.5,Iris-versicolor
...
6.3,3.3,6.0,2.5,Iris-virginica
5.8,2.7,5.1,1.9,Iris-virginica
7.1,3.0,5.9,2.1,Iris-virginica
...
```



Regression

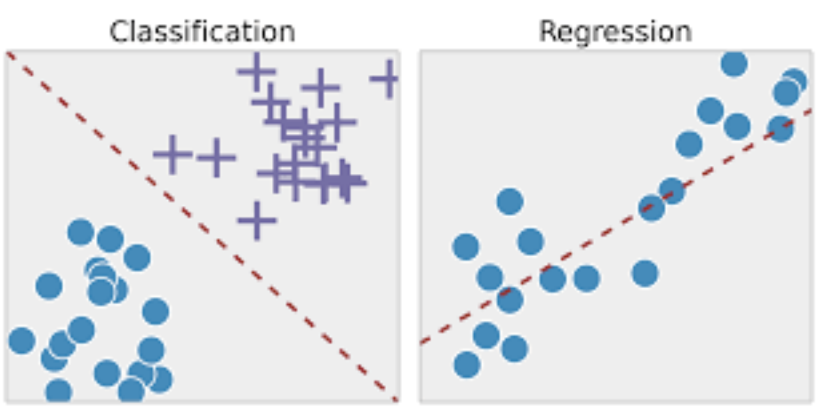
- The data is represented in some vector space.
- Let's say we have input data.

$$x = (x_1, x_2, x_3, \ldots)$$

- For each of the input data, it's represented by a vector. We call it features.
- For each of the input data, we know their corresponding value.

$$y = (y_1, y_2, y_3, \ldots)$$

- The goal is to give the appropriate value for some given new data.



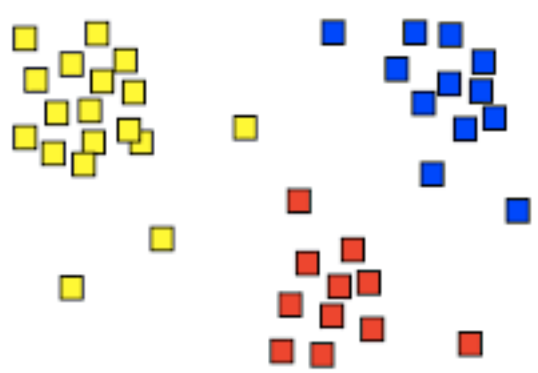
Example:
- Stock price prediction?
- Quality control?

Clustering

- The data is represented in some vector space.
- Let's say we have input data.

$$x = (x_1, x_2, x_3, \ldots)$$

- For each of the input data, it's represented by a vector. We call it features.
- We want to group the data into clusters so that data points in the same cluster are similar to each other than data points belonging to different clusters.



Example:

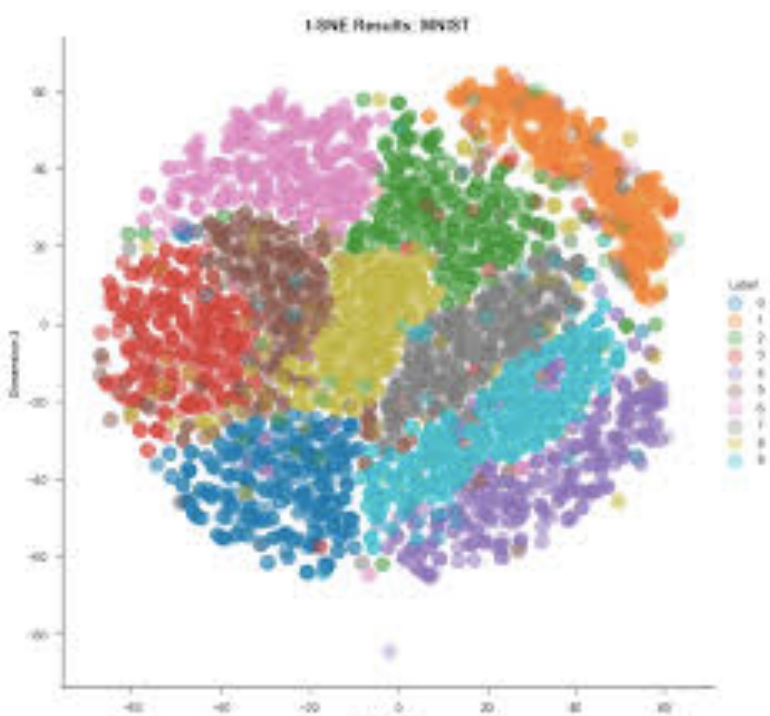
- Stories into topics.
- Movies into categories.
- Genes by similar function.

Dimensionality Reduction

- The data is represented in some vector space.
- Let's say we have input data.

$$x = (x_1, x_2, x_3, \ldots)$$

- For each of the input data, it's represented by a vector. We call it features.
- We want to discover and model the intrinsic dimension of the data and/or to project the high dimensional data into lower dimension while preserve the relevant information.



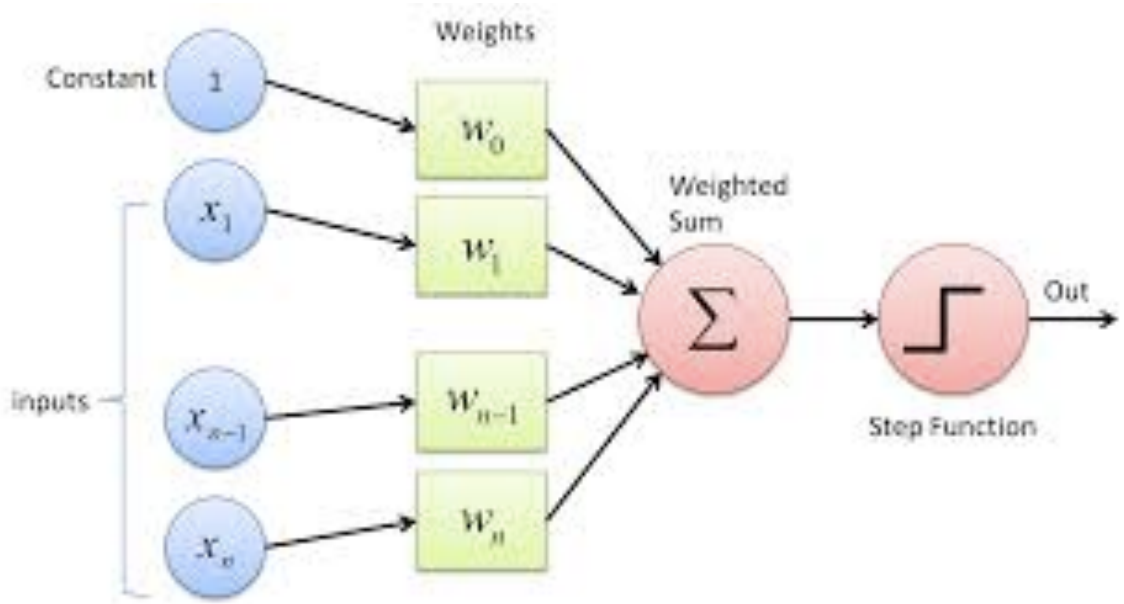
Example:

- t-sne/umap for visualization.

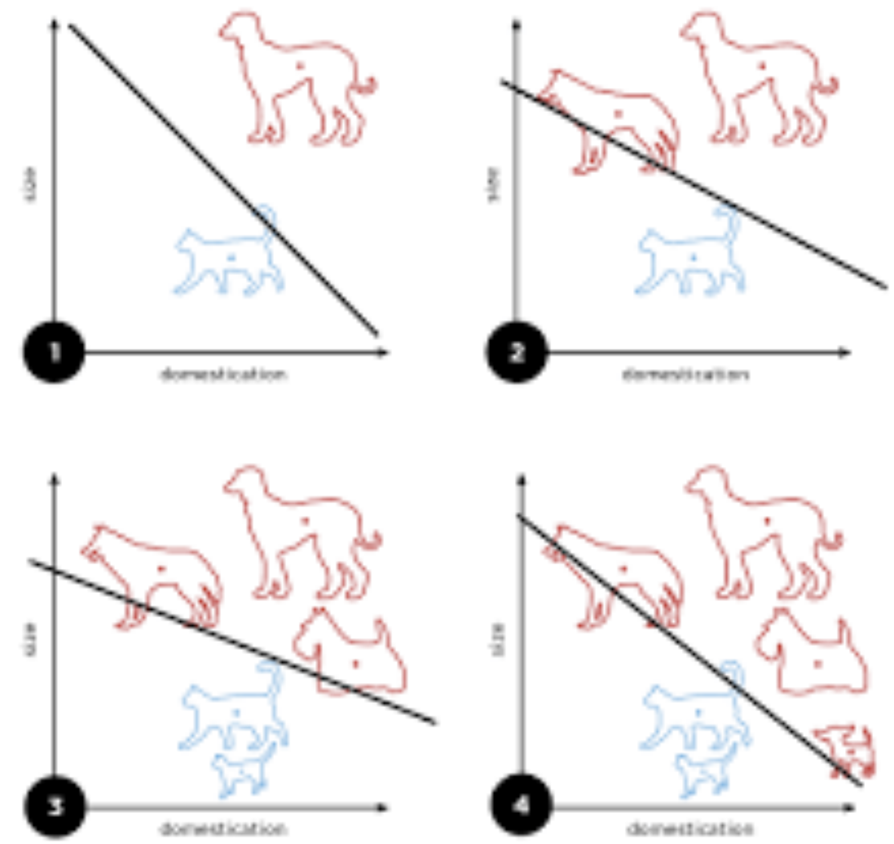
Artificial Neural Network

- Perceptron
- Multiple Layer Perceptron (MLP)
- Stacked Boltzmann Machine
- Back-propagation
- Convolutional Neural Network
- Information Bottleneck and Autoencoder
- Recurrent Neural Network
- Go Deeper!
- …

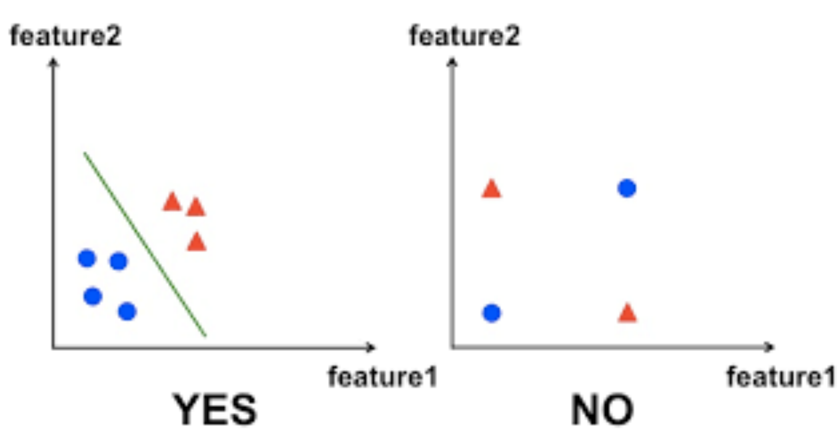
Perceptron



Weights
Constant
1
w_0
x_1
w_1
Weighted
Sum
Σ
Out
inputs
x_{n-1}
w_{n-1}
Step Function
x_n
w_n

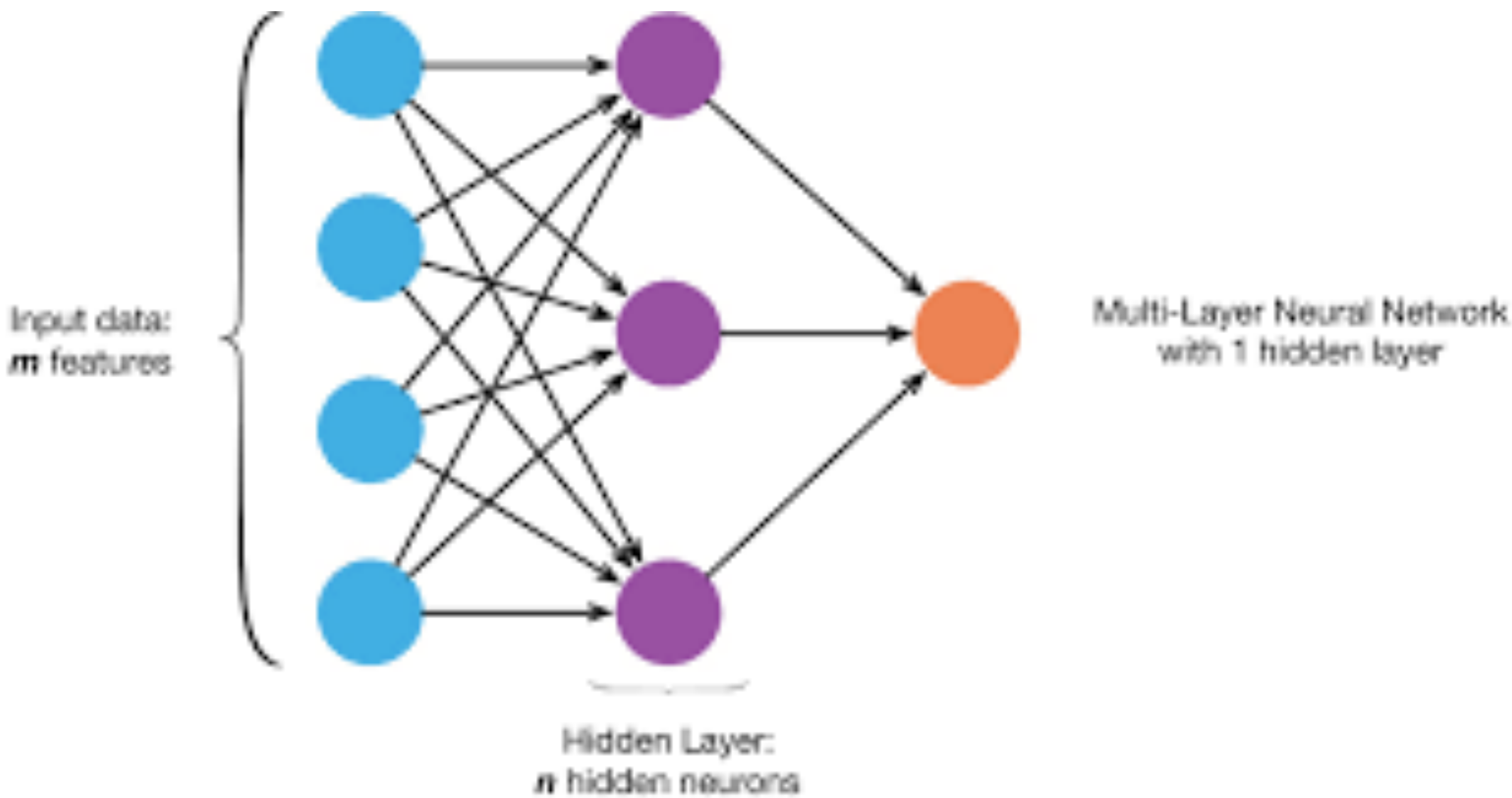


1
2
3
4

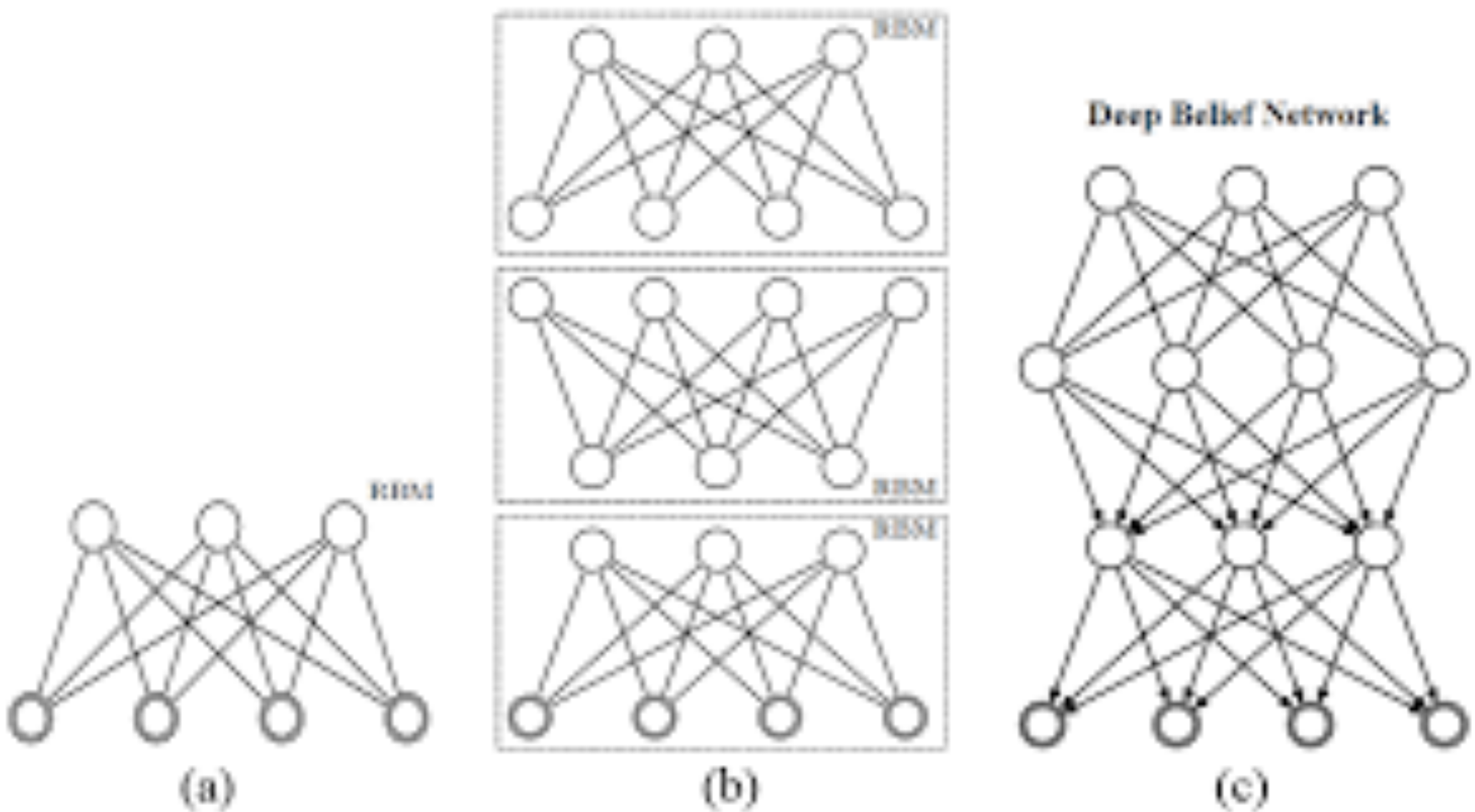


feature2
feature1
YES
feature2
feature1
NO

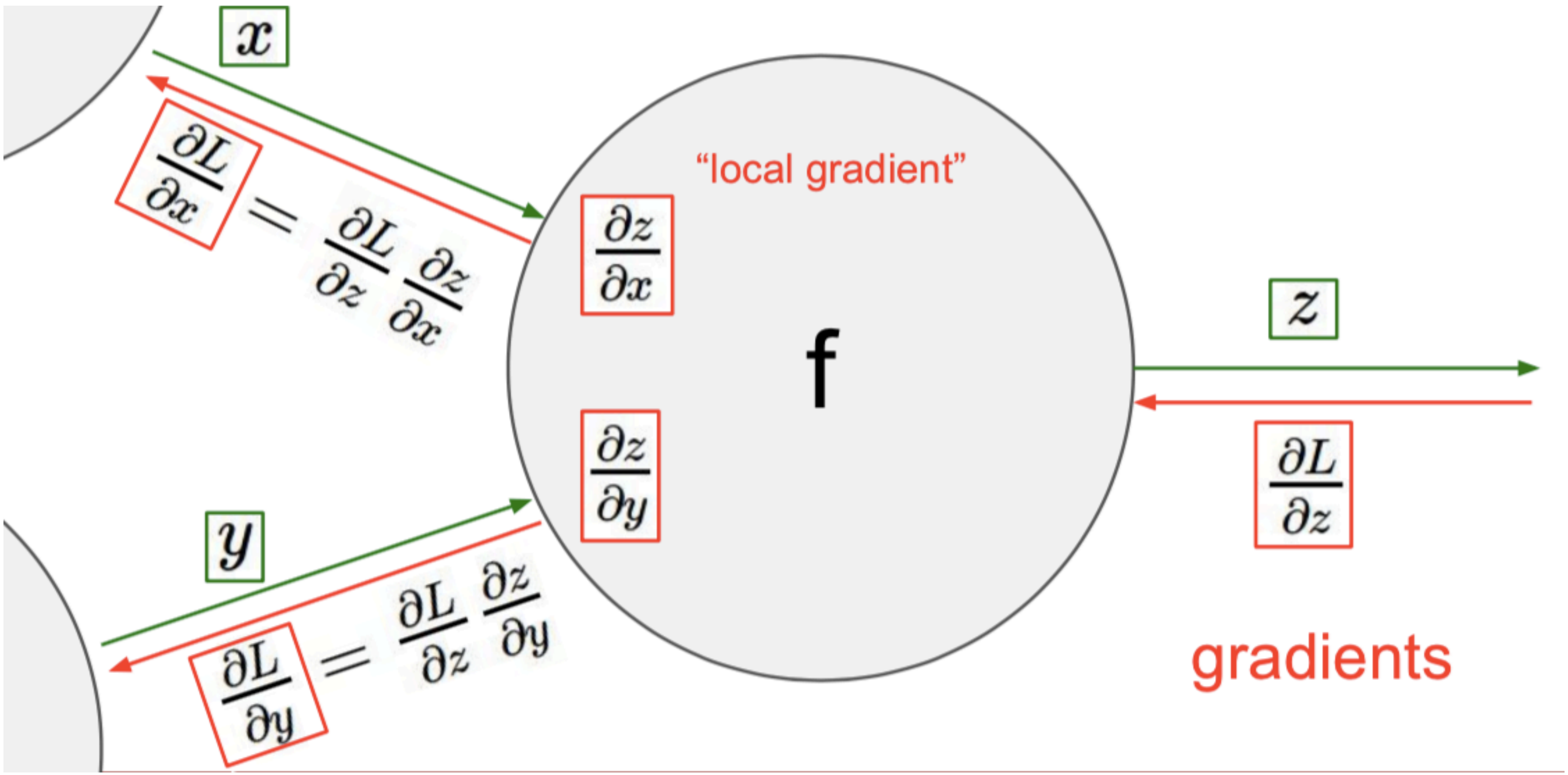
Multiple Layer Perceptron



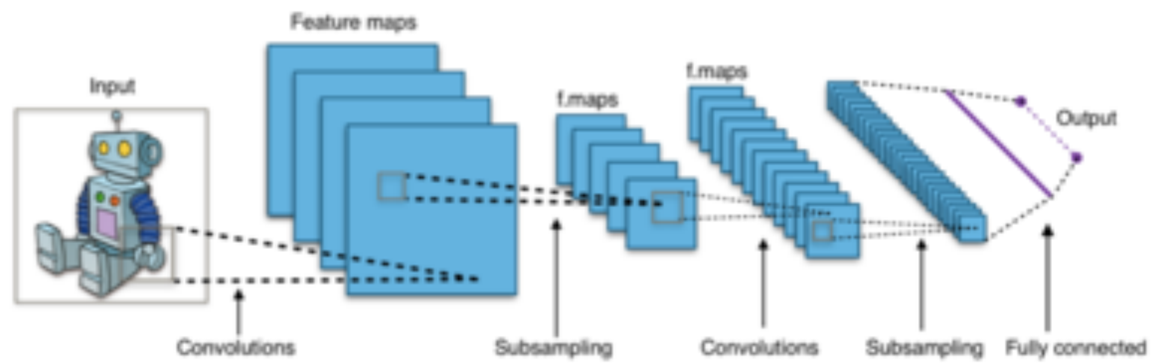
Stacked Restricted Boltzmann Machine



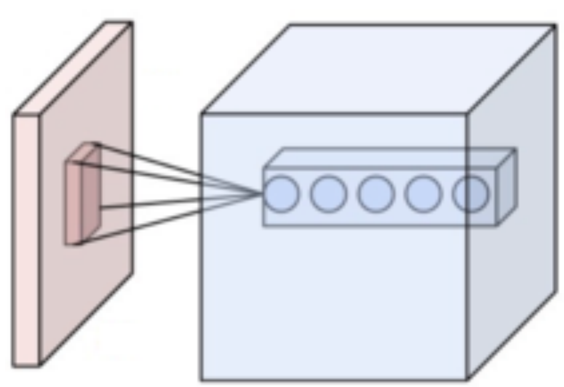
Back-propagation



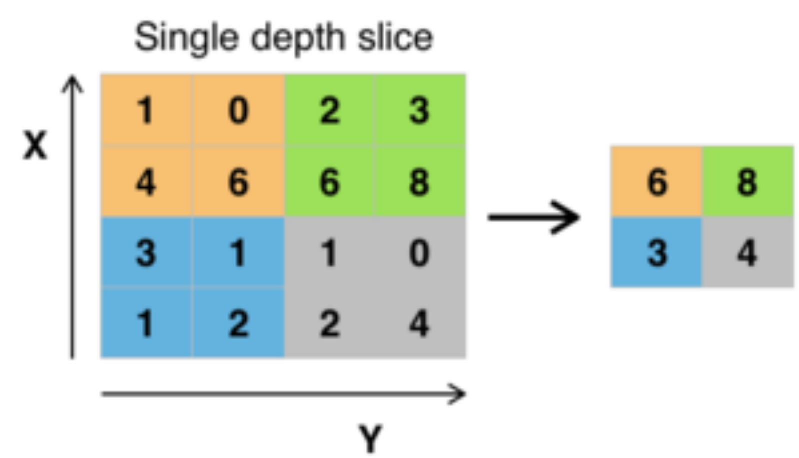
Convolutional Neural Network



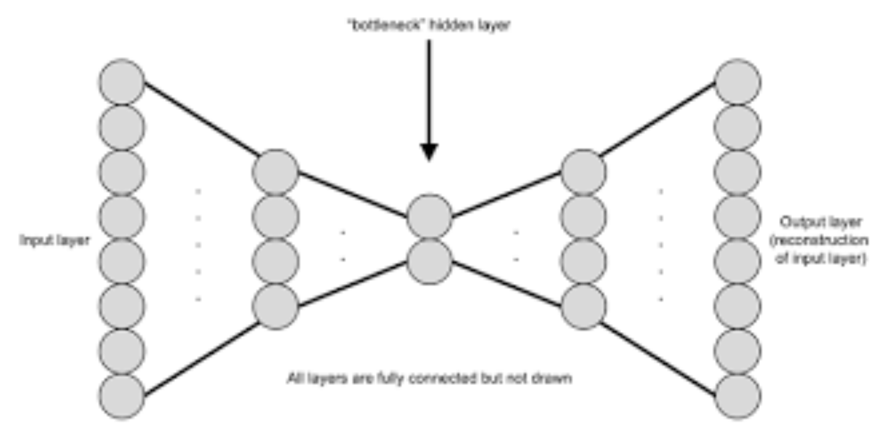
Convolutional Operator

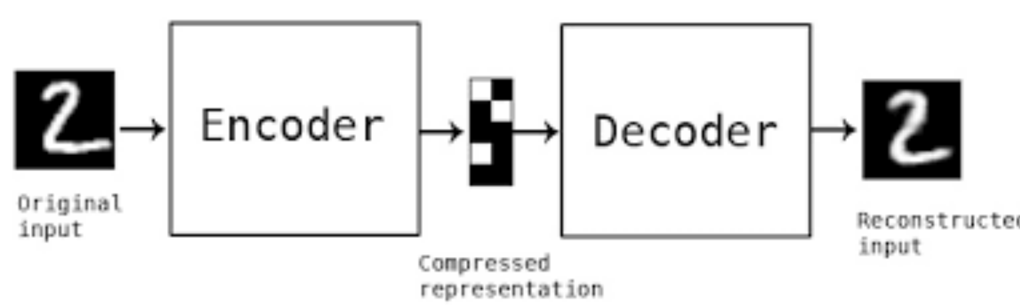


Pooling

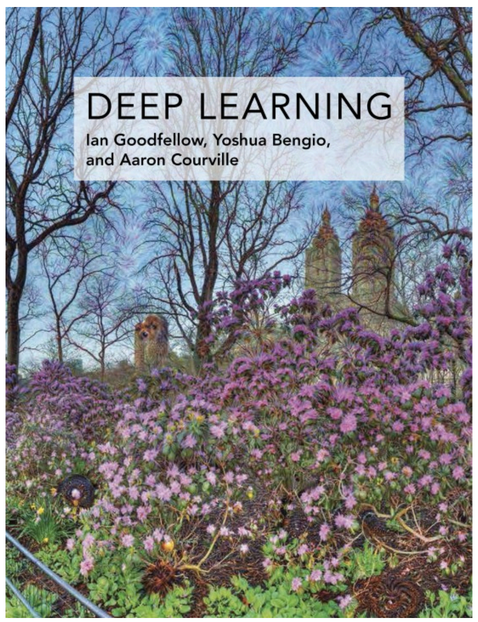


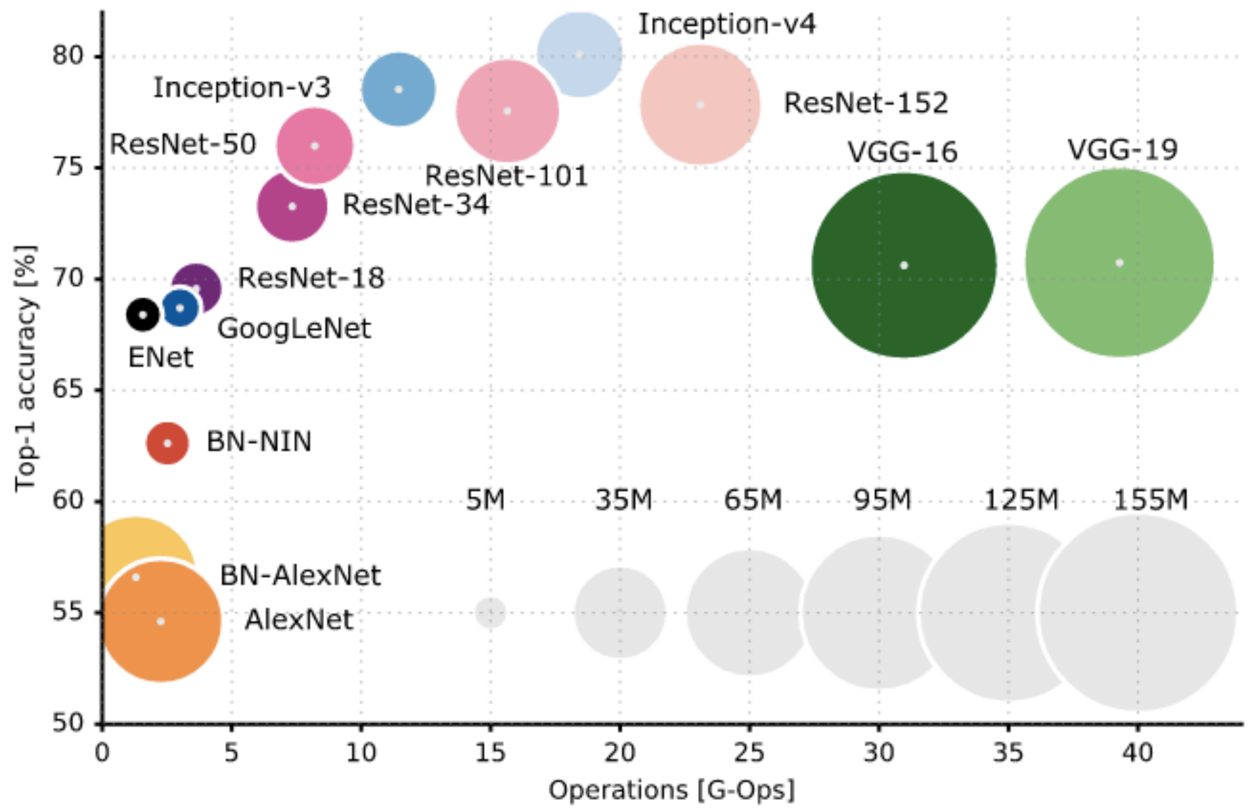
Information Bottleneck & Autoencoder

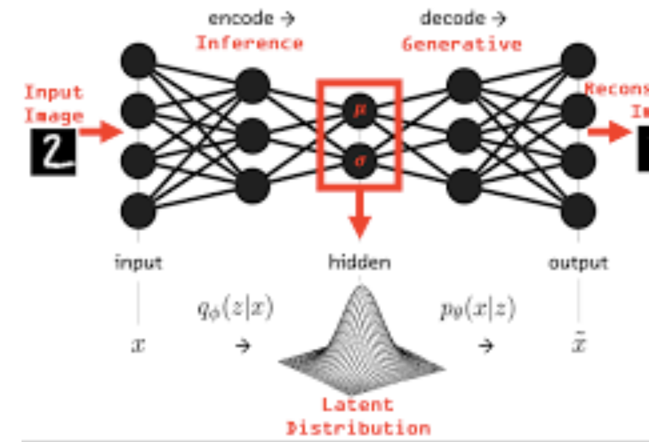


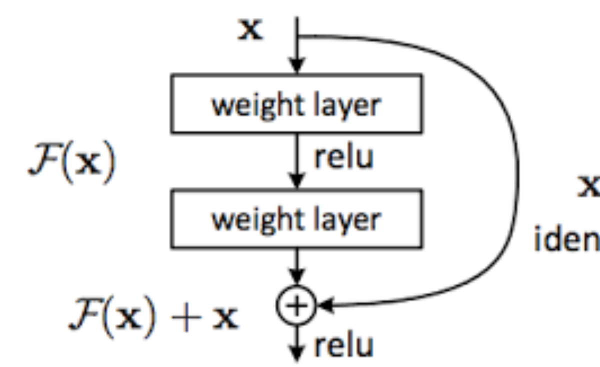


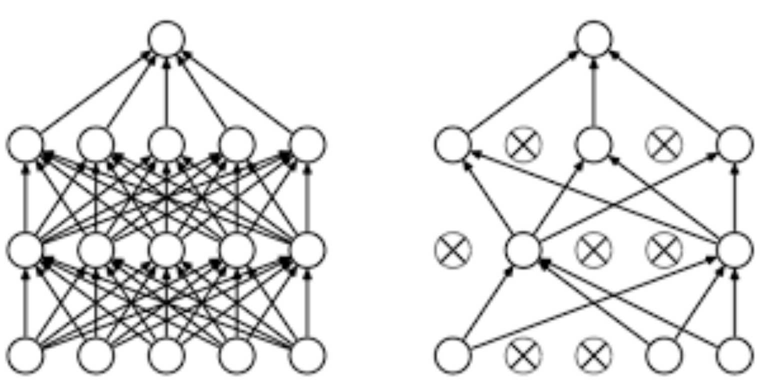
Deep Networks

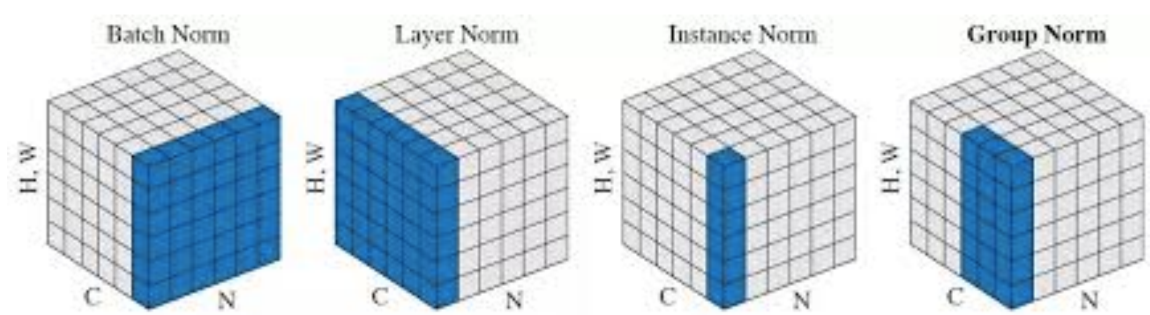


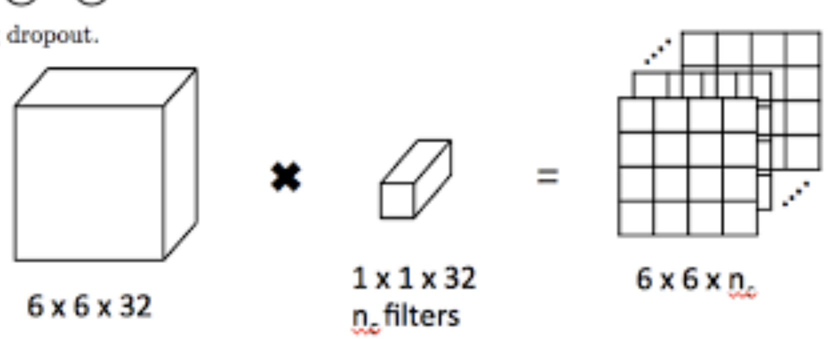


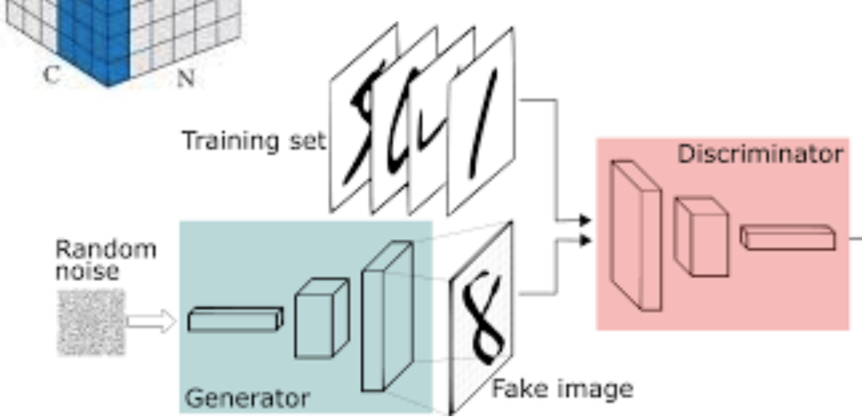




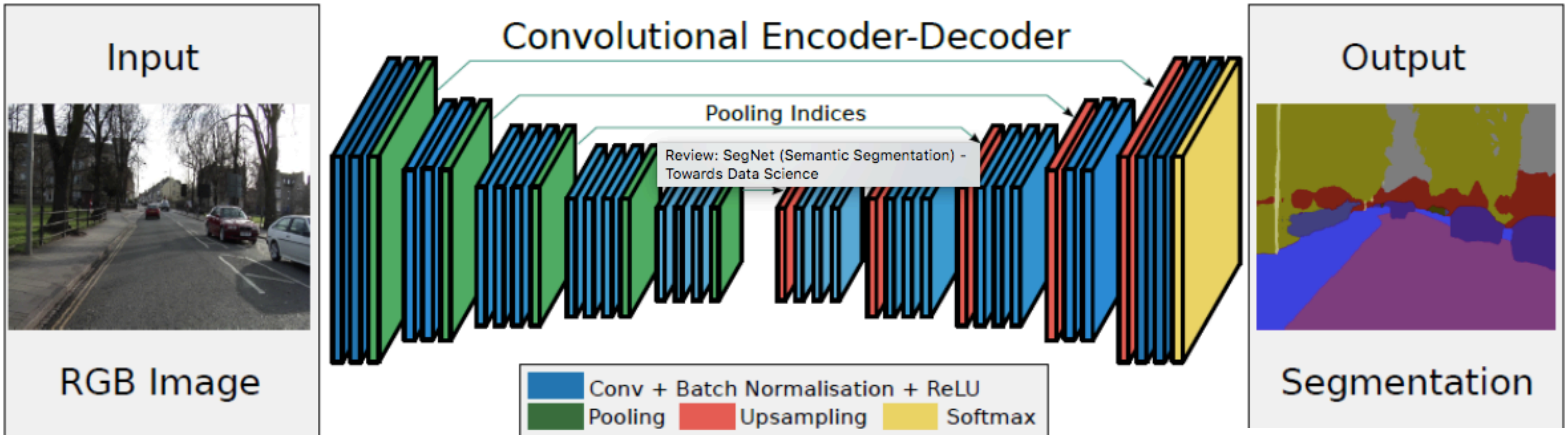

(a) Standard Neural Net (b) After applying dropout.



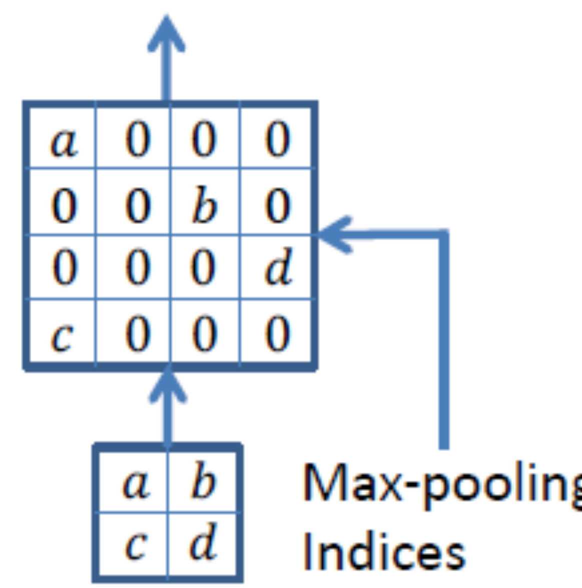




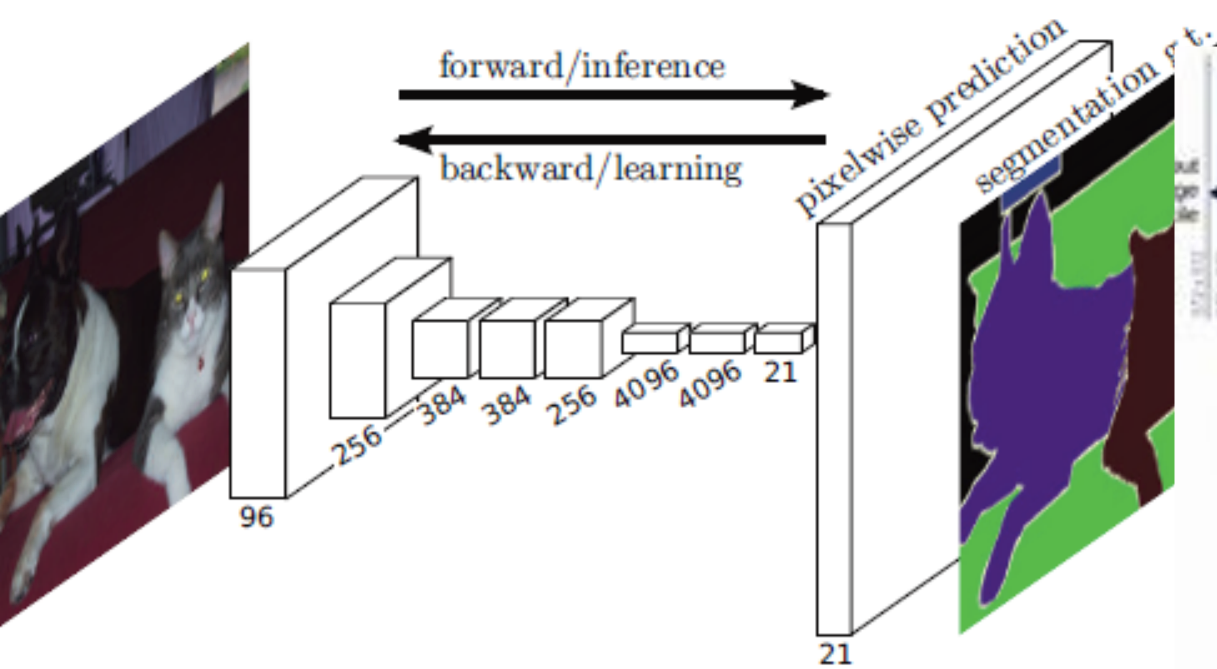
Segmentation

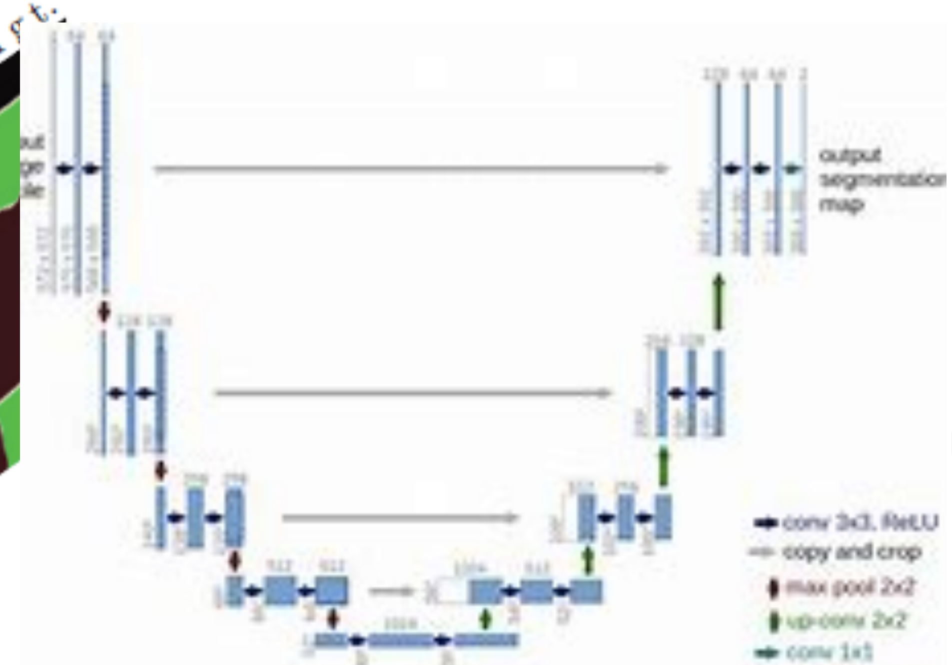


SegNet: Encoder Decoder Architecture



Upsampling Using Max-Pooling Indic

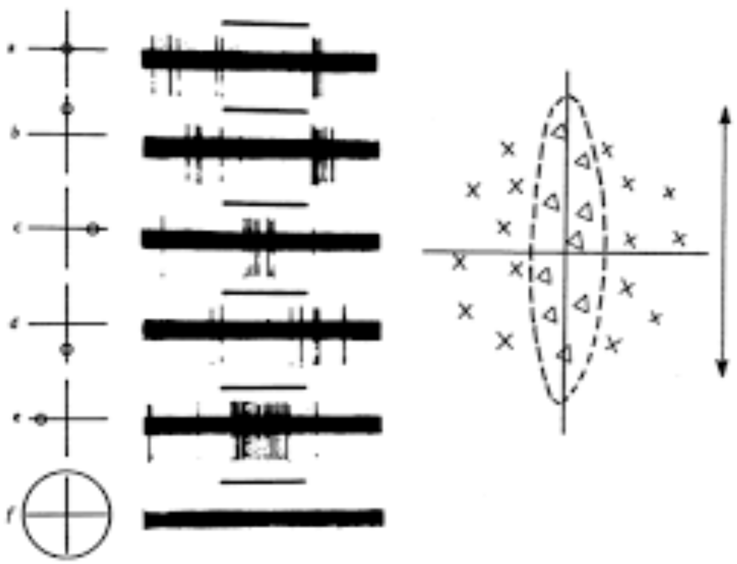


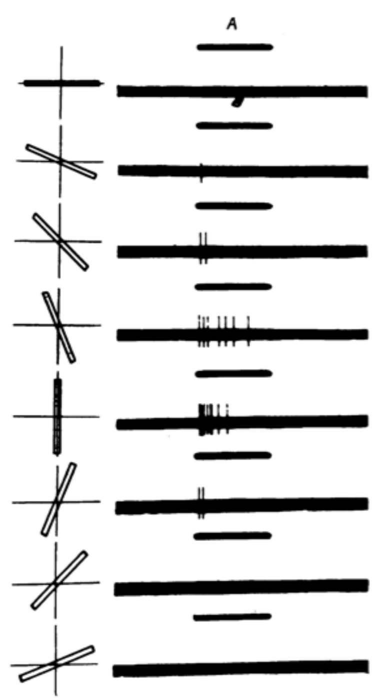


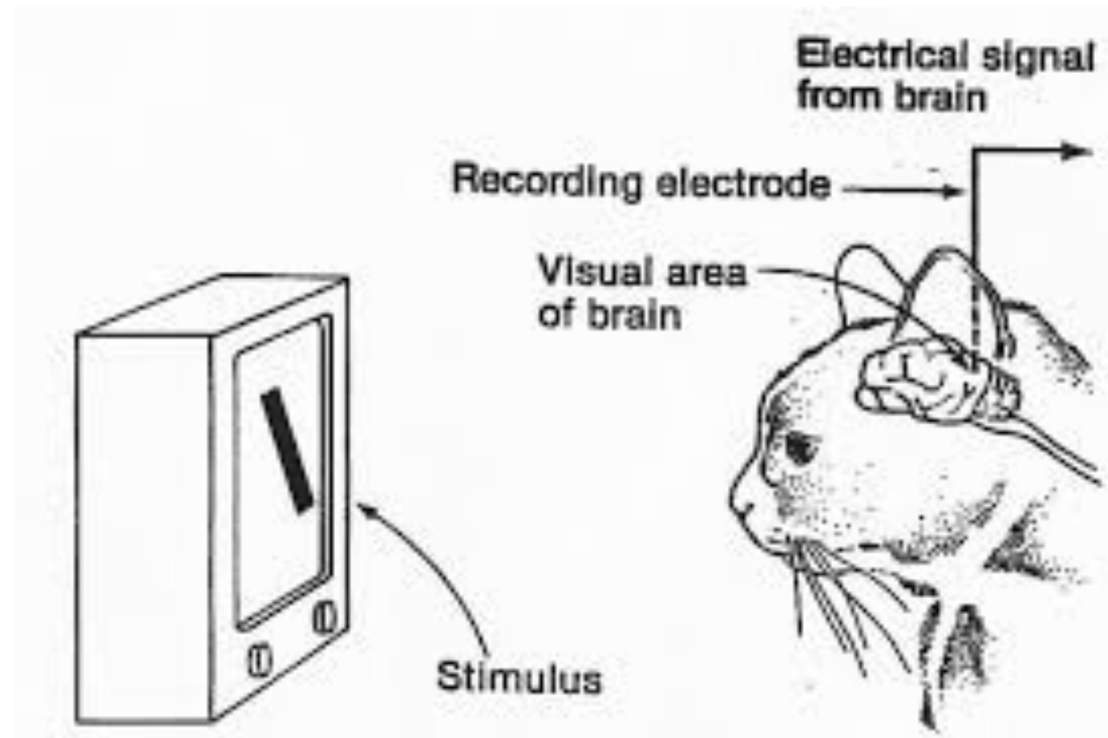
Feature Map / Filter Number Along Layers

Some neuroscience flavor

Vision System and Convolutional Neural Network

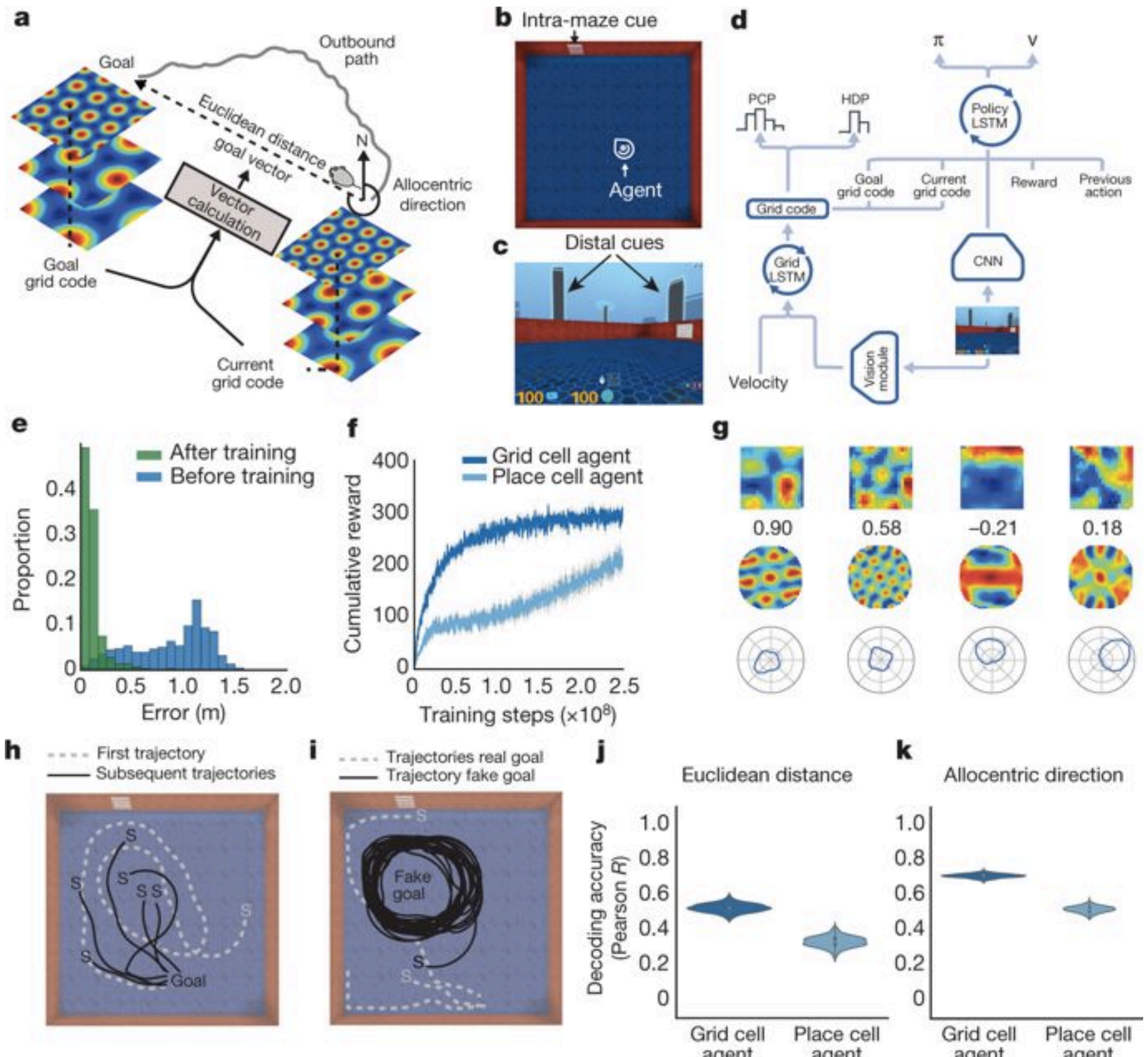


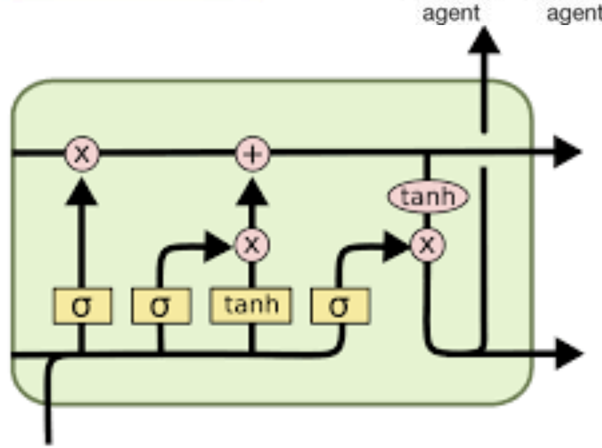


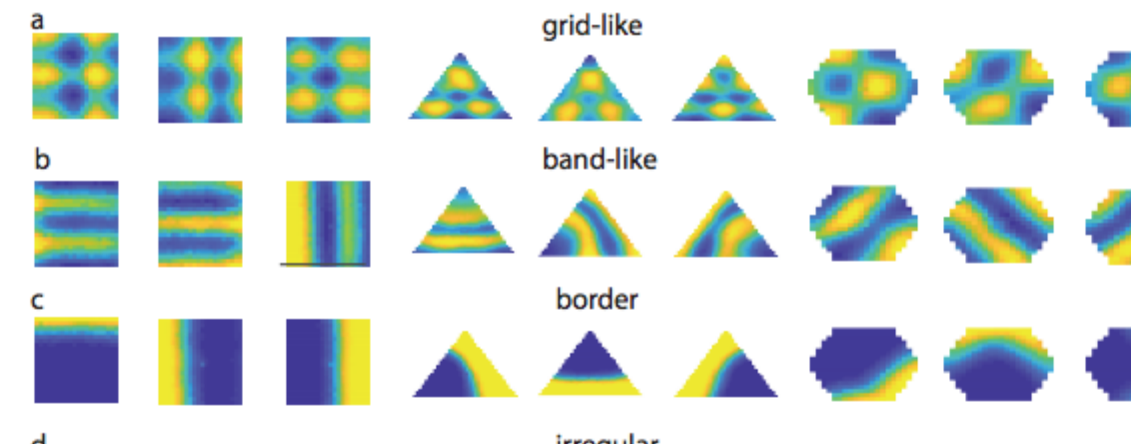


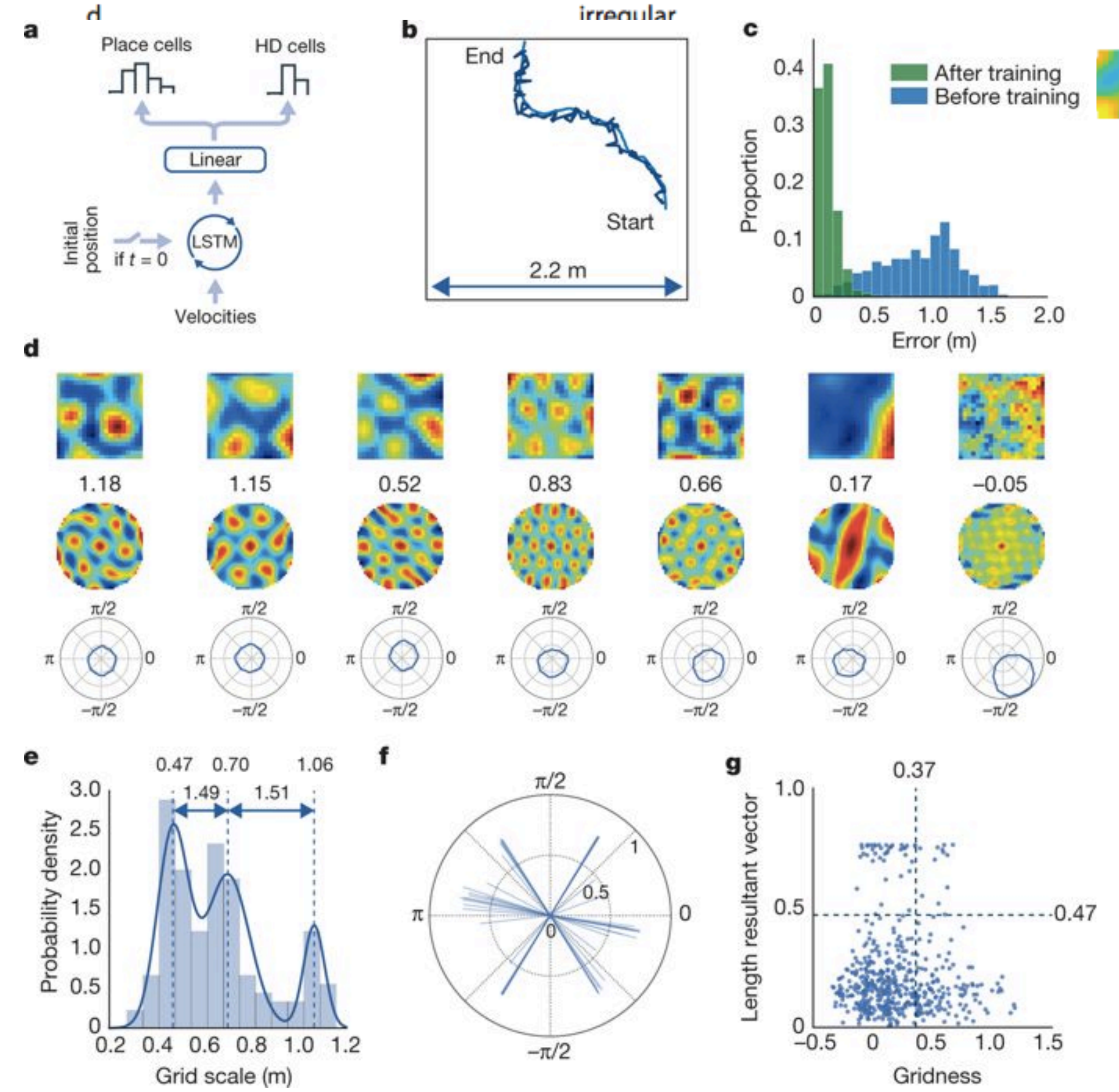
- **Hubel, Wiesel, and Sperry**
- **Nobel Prize 1982**

Grid Cell and LSTM









Summary

- Machine learning concept and boundary.
- Applications of machine learning.
- 3 Types of machine learning.
- Commonly seen machine learning problem.
- Artificial neural networks/deep learning.
- Examples of join work on neuroscience and ANN.

Review (probability)

- Random Variable/Expectation/Variance
- Conjugate prior/Exponential family/sufficient statistic
- Bayesian

$$f_X(x \mid \boldsymbol{\theta}) = h(x) \exp\Big(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \mathbf{T}(x) - A(\boldsymbol{\theta})\Big)$$

- Bias-variance decomposition

$$\mathrm{E}\Big[\big(y - \hat{f}(x)\big)^2\Big] = \Big(\mathrm{Bias}\big[\hat{f}(x)\big]\Big)^2 + \mathrm{Var}\big[\hat{f}(x)\big] + \sigma^2$$

where

$$\mathrm{Bias}\big[\hat{f}(x)\big] = \mathrm{E}\big[\hat{f}(x)\big] - f(x)$$

and

$$\mathrm{Var}\big[\hat{f}(x)\big] = \mathrm{E}[\hat{f}(x)^2] - \mathrm{E}[\hat{f}(x)]^2$$

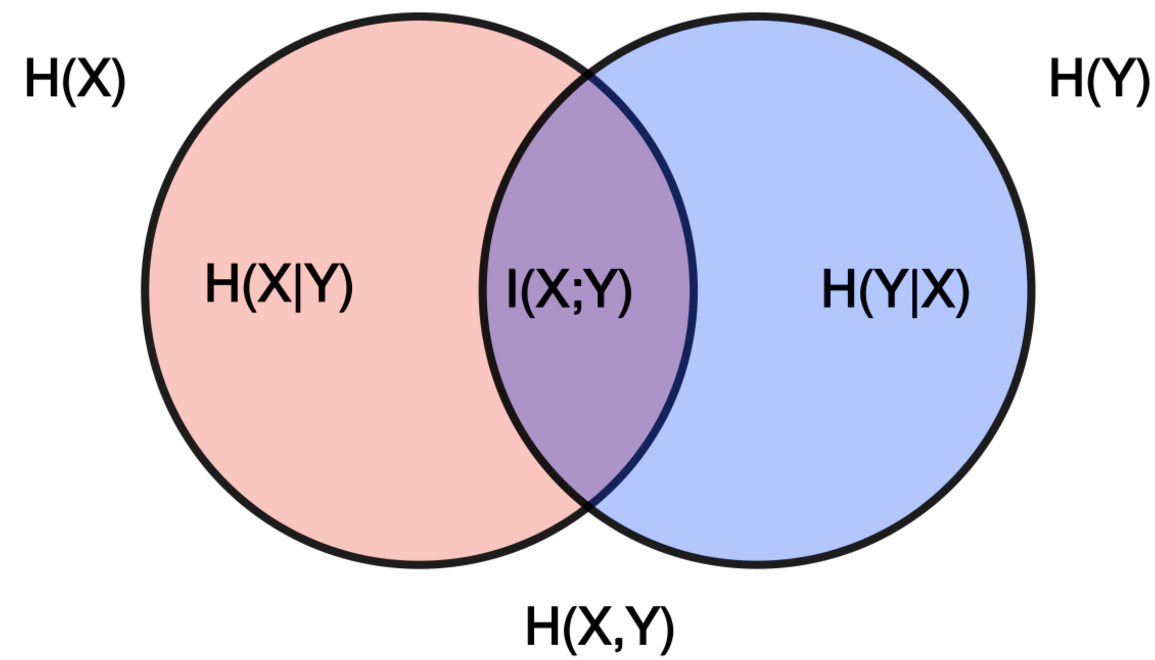
$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Review (information theory)

- Entropy

$$S = -\sum_i P_i \log P_i$$

- Cross entropy
- Mutual information



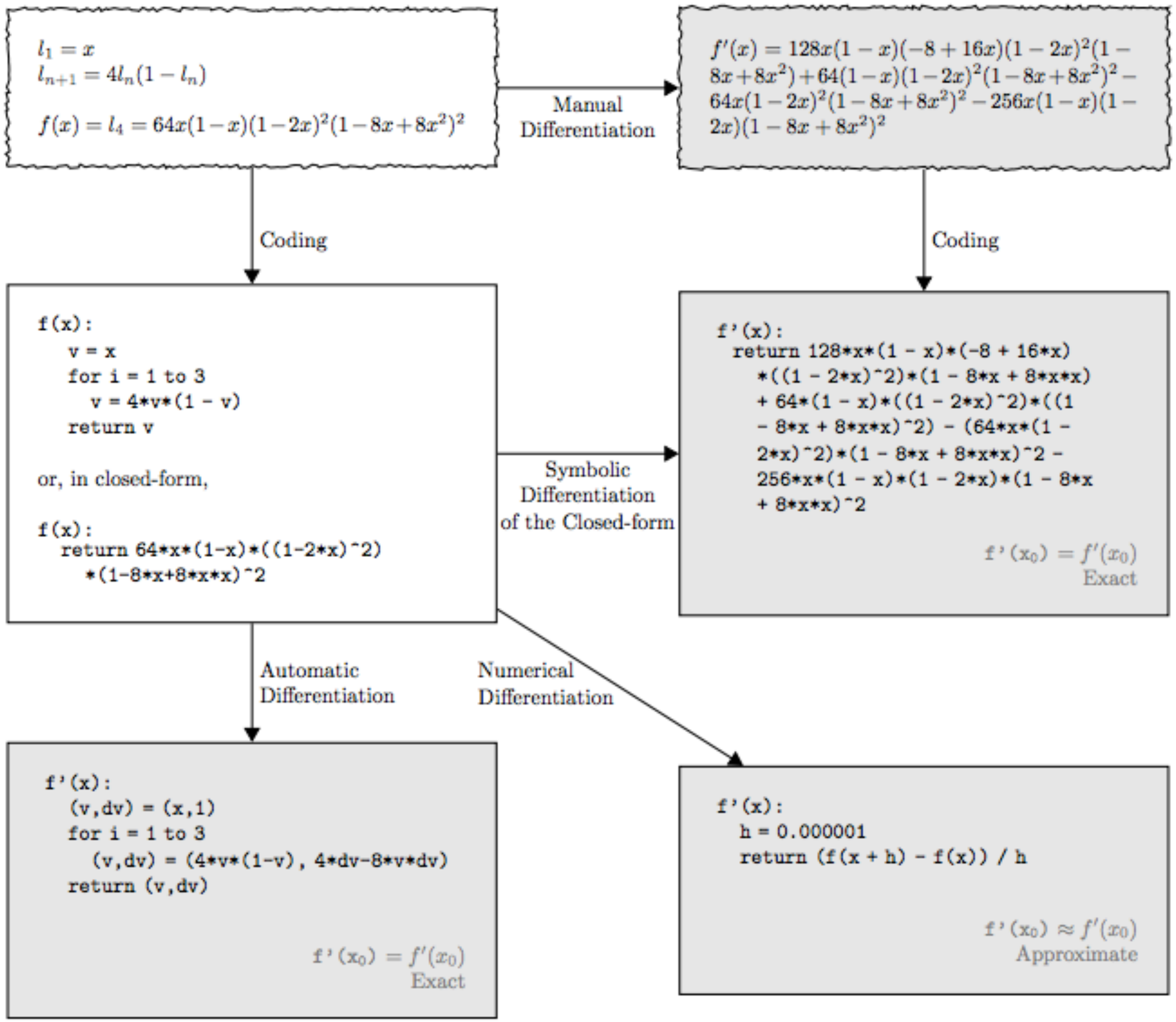
Reading

- C. Bishop, Pattern recognition and machine learning(PRML), chapter 5 Neural networks.

homework assignment

- PRML, 5.1, 5.4
- In a machine learning competition on a binary classification problem, you are allowed to submit your answer and the competition will give you a number in [0,1] back indicating the correct rate. How many submissions you need in order to crack out all the labels? Say there are 1024 samples.

Computation of derivatives



- **Manually working out derivatives and coding them;**
- **numerical differentiation;**
- **symbolic differentiation;**
- **automatic differentiation/ algorithmic differentiation.**

Automatic Differentiation in Machine Learning: a Survey, 2018 https://www.youtube.com/watch?v=qGvB0qno9OY