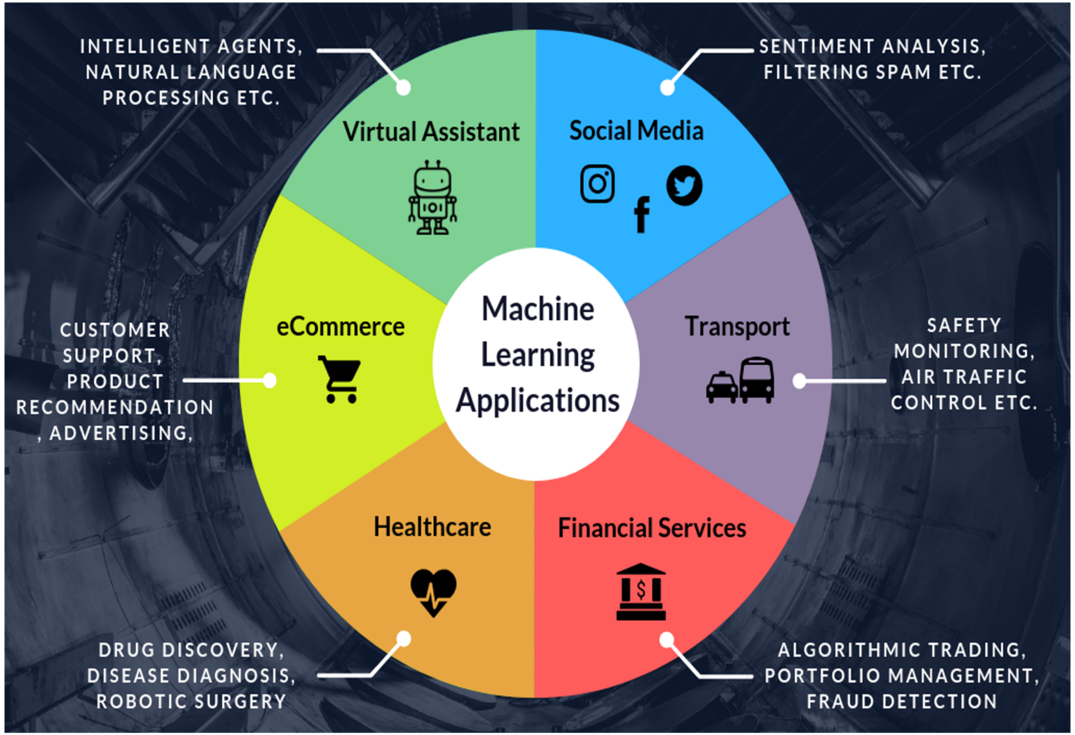
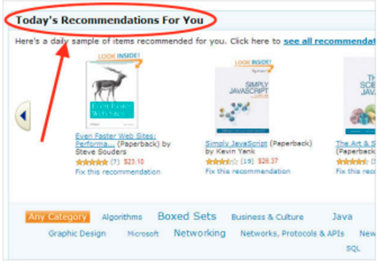
Applications of Machine Learning



https://www.quantinsti.com/blog/machine-learning-basics

Applications of Machine Learning: Examples

spam filtering



recommendation systems

fraud detection

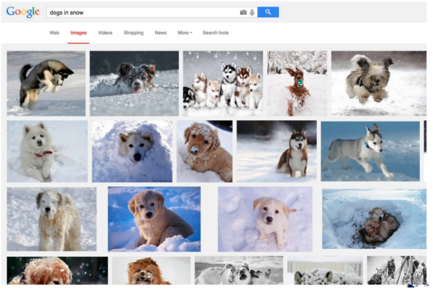


Image retrieval

self-driving cars

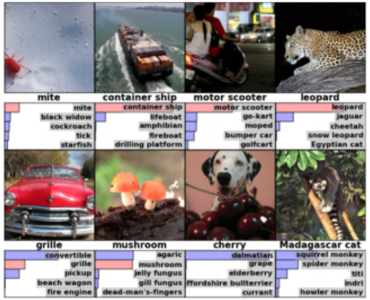
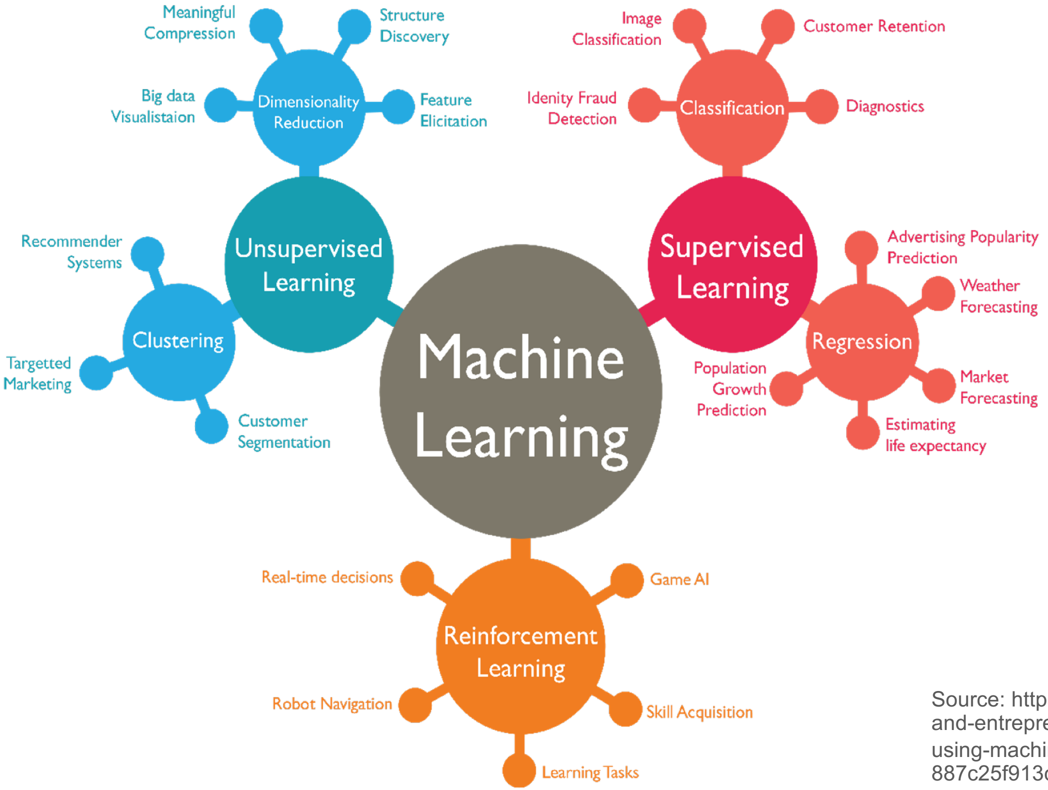


image recognition

stock market analysis

Types of Machine Learning



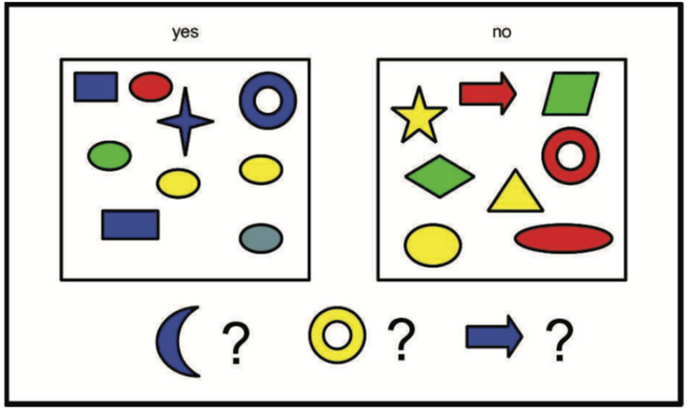
Source: https://medium.com/marketing-and-entrepreneurship/10-companies-using-machine-learning-in-cool-ways-887c25f913c3

Supervised Learning

- Given a labeled set of input-output pairs $D = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$
  - D is the training set
  - N is the number of labeled examples
  - $\boldsymbol{x}$ is a d-dimensional vector of features/attributes/covariates
  - y is a response variable
- Learn a function $y = f(\boldsymbol{x})$
  - Classification when y is categorical, $y \in \{1, \dots, C\}$
  - Regression when y is real-valued, $y \in \mathbb{R}$
- Probabilistic modeling $p(y|\boldsymbol{x}, D, \theta)$
  - θ denotes the set of model parameters

Classification

- When the response variable is categorical $y \in \{1, \ldots, C\}$
  - When $C = 2$, binary classification
  - When $C > 2$, multiclass classification
  - When $\boldsymbol{x}_i$ may have multiple labels, multi-label classification



(a)

D features (attributes)

N cases

Color	Shape	Size (cm)
Blue	Square	10
Red	Ellipse	2.4
Red	Ellipse	20.7

Label
1
1
0

(b)

Classification: Classifying Flowers

Three types of iris flowers: setosa, versicolor and virginica.

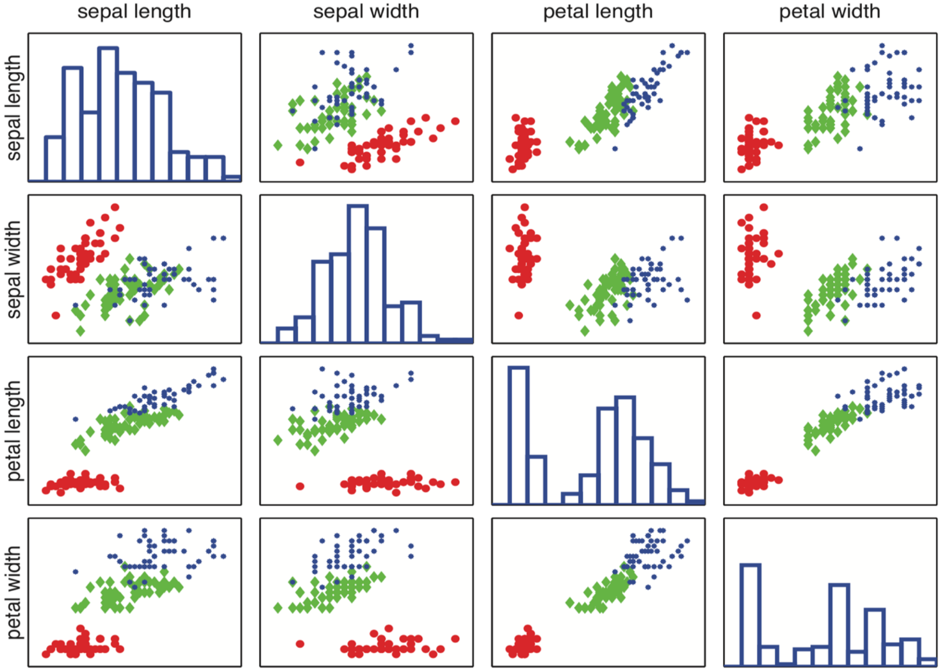


Figure 1.4 Visualization of the Iris data as a pairwise scatter plot. The diagonal plots the marginal histograms of the 4 features. The off diagonals contain scatterplots of all possible pairs of features. Red circle = setosa, green diamond = versicolor, blue star = virginica.

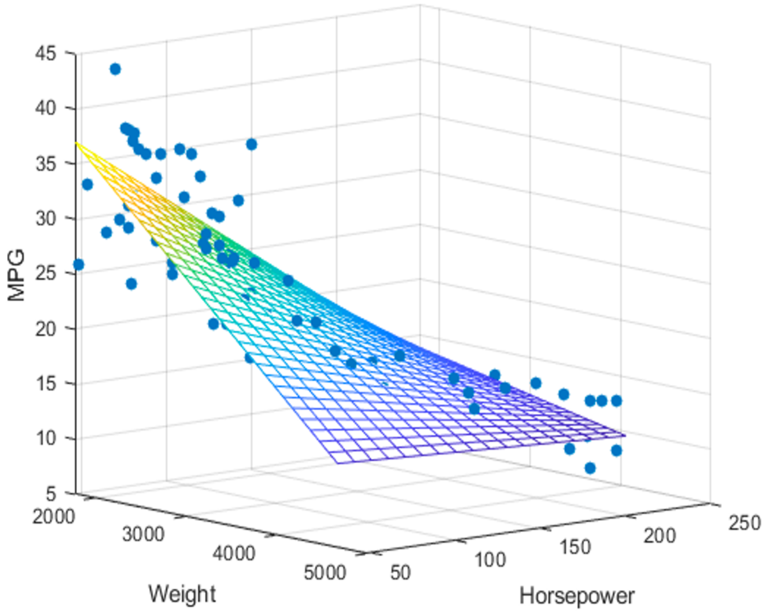
Classification: Emotion Detection

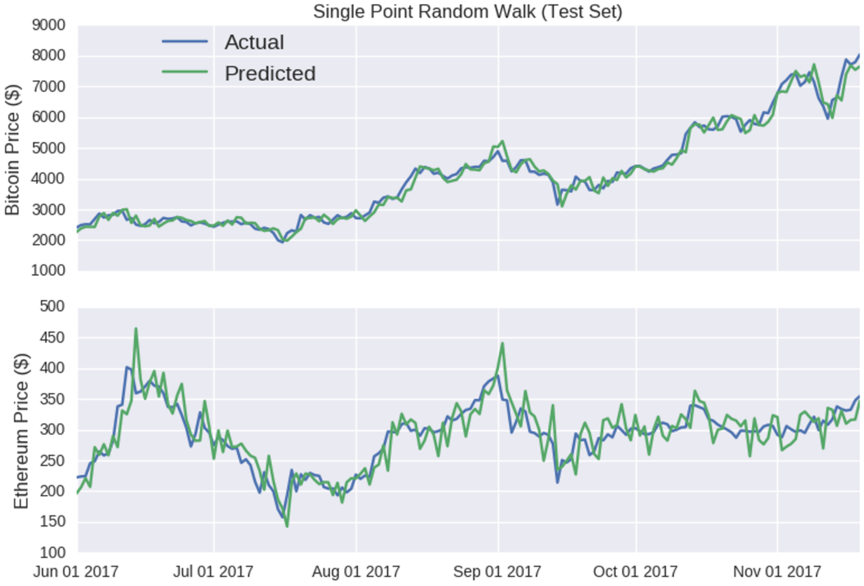
- Anger
- Disgust
- Fear
- Happiness
- Neutral
- Sadness
- Surprise

https://www.freecodecamp.org/news/facial-emotion-recognition-develop-a-c-n-n-and-break-into-kaggle-top-10-f618c024faa7/

Regression

- When the response variable is real-valued $y \in \mathbb{R}$



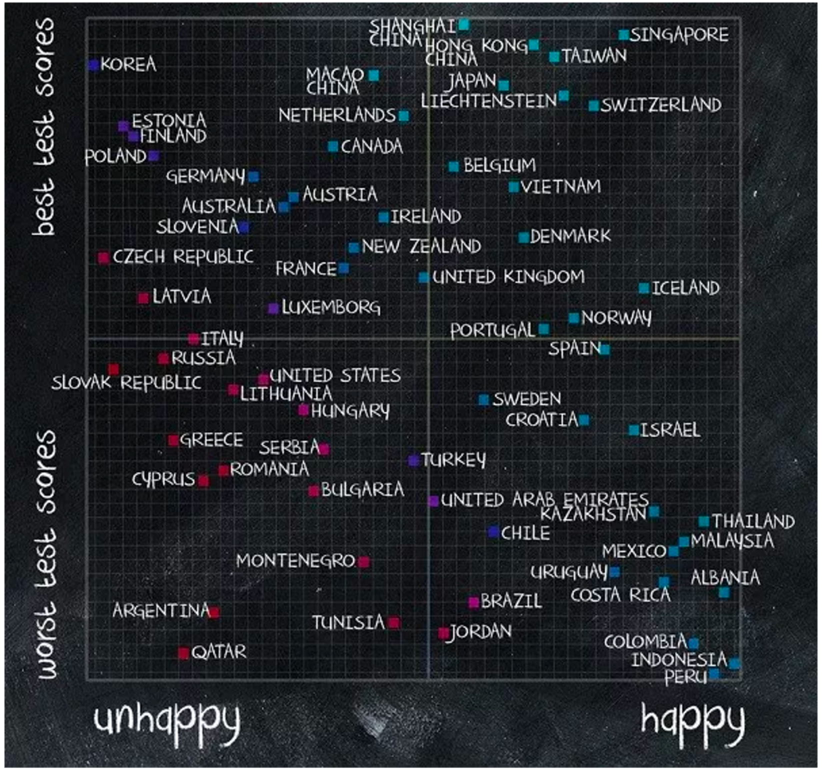


Unsupervised Learning

- Given a set of inputs $D = \{\boldsymbol{x}_i\}_{i=1}^{N}$
  - D is the input data
  - N is the number of examples
  - $\boldsymbol{x}$ is a d-dimensional vector of features/attributes/covariates
  - there is <u>no</u> response variable
- Learn a density function $p(\boldsymbol{x}_i|\theta)$
  - θ denotes the set of model parameters

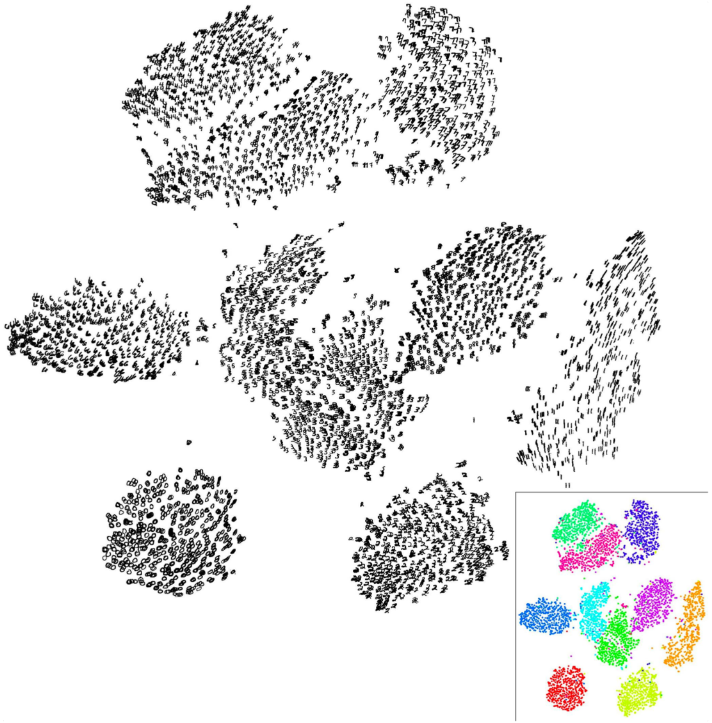
Clustering

- Clusters:
  - subgroups or subpopulations in the data
- Goals:
  - Discovering the subgroups
  - Estimating which subgroup a data point belongs to



https://i2.wp.com/techielobang.com/blog/wp-content/uploads/2014/01/best-school-happiest-kids.jpg

Clustering: MNIST Handwritten Digits



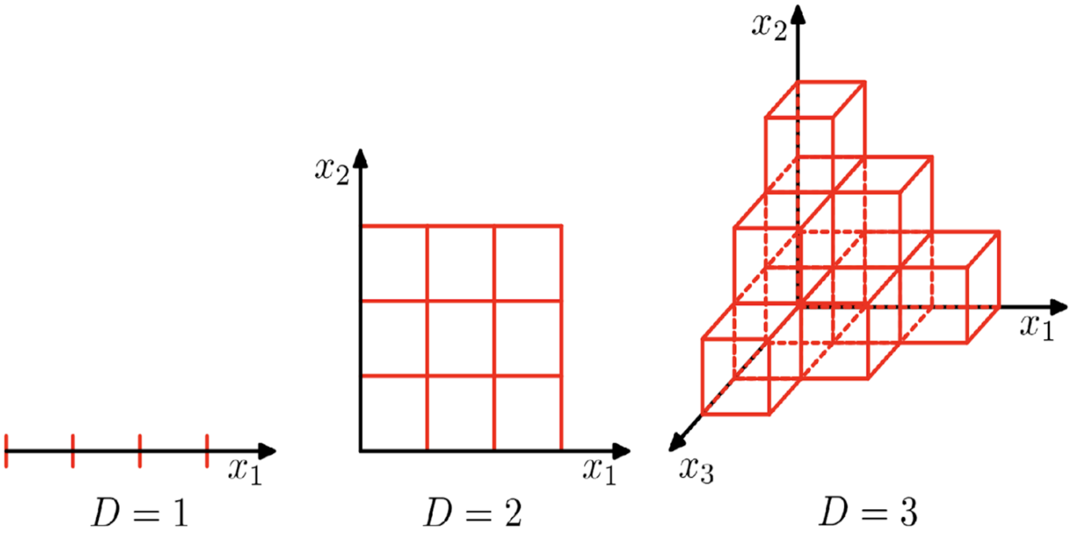
Source: http://lvdmaaten.github.io/tsne/examples/mnist_tsne.jpg

"The original black and white (bilevel) images from NIST were size normalized to fit in a 20x20 pixel box while preserving their aspect ratio. The resulting images contain grey levels as a result of the anti-aliasing technique used by the normalization algorithm. the images were centered in a 28x28 image by computing the center of mass of the pixels, and translating the image so as to position this point at the center of the 28x28 field."

Source: http://yann.lecun.com/exdb/mnist/

Curse of Dimensionality

Figure 1.21 Illustration of the curse of dimensionality, showing how the number of regions of a regular grid grows exponentially with the dimensionality D of the space. For clarity, only a subset of the cubical regions are shown for $D = 3$.



Occam's Razor Principle

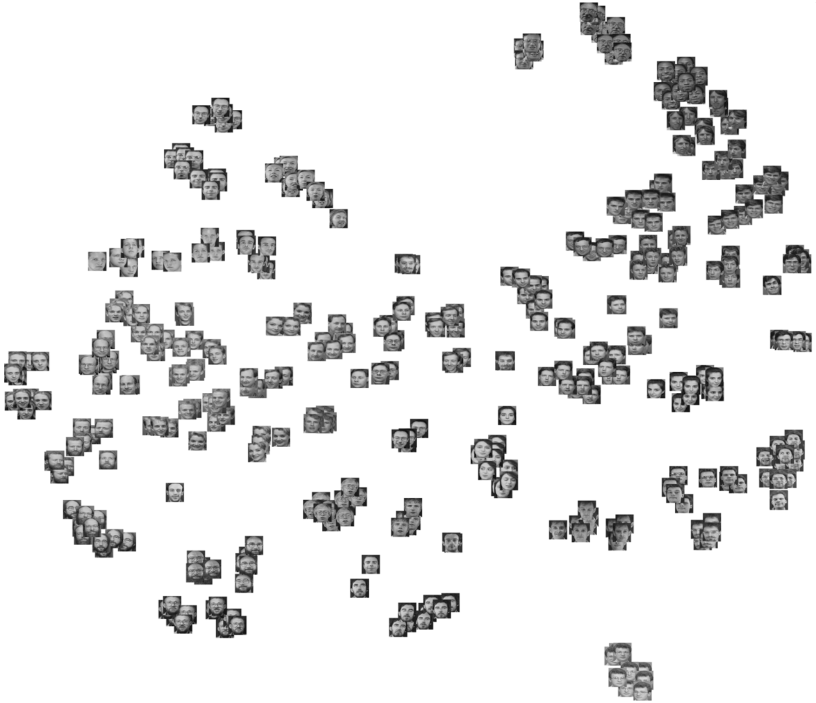
Occam's Razor is a principle that likes simplicity. It says that the simplest solution is usually the best one. In machine learning, this means that if we have two models that work about as well as each other, we should choose the simpler one.

Dimensionality Reduction

Source:
https://www.flickr.com/photos/digitalanthill/4617998281/

Projecting the data to a lower-dimensional subspace that captures the "essence" of the data

Dimensionality Reduction: Olivetti Faces



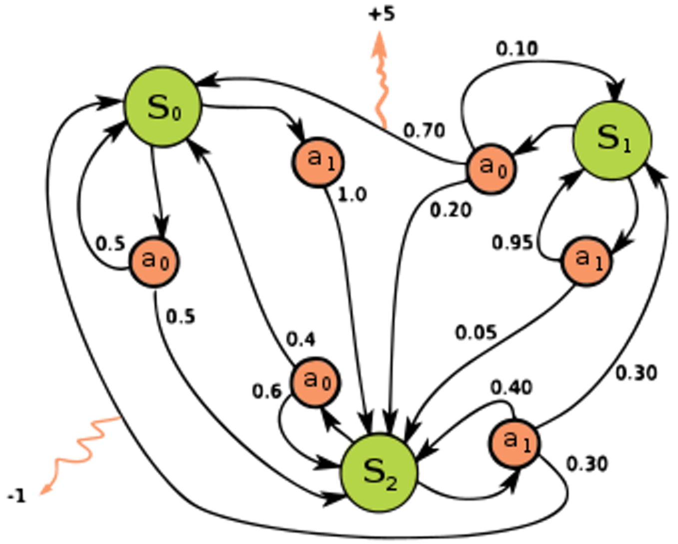
Source:
http://lvdmaaten.github.io/tsne/examples/olivetti_tsne.jpg

”There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).”

Source:
https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

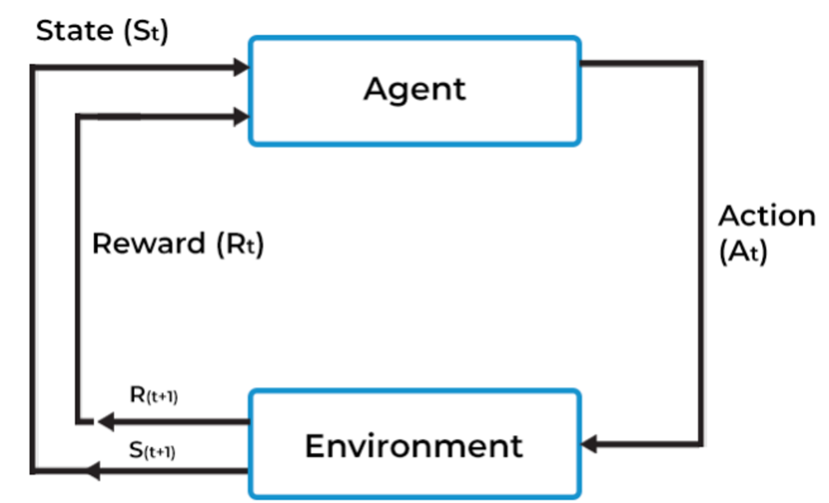
Markov Decision Process

Markov decision process (**MDP**) is a discrete-time stochastic control process. It provides a mathematical framework for modeling decision making in situations where outcomes are partly random and partly under the control of a decision maker



Source: https://en.wikipedia.org/wiki/Markov_decision_process

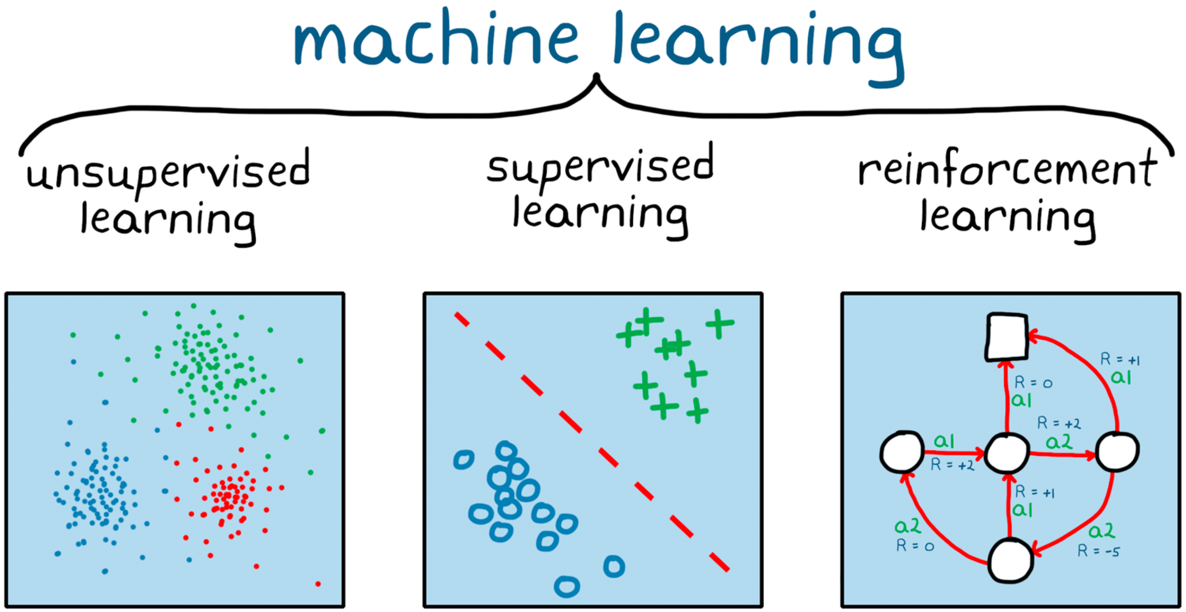
Reinforcement Learning



Reinforcement learning (**RL**) is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward

Source: https://en.wikipedia.org/wiki/Reinforcement_learning

Reinforcement Learning vs Traditional Learning



Source: https://www.mathworks.com/discovery/reinforcement-learning.html

Summary

- Introduction to Machine Learning
- Statistics vs Machine Learning
- Machine Learning vs Deep Learning
- Applications of Machine Learning
- Types of Machine Learning
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning