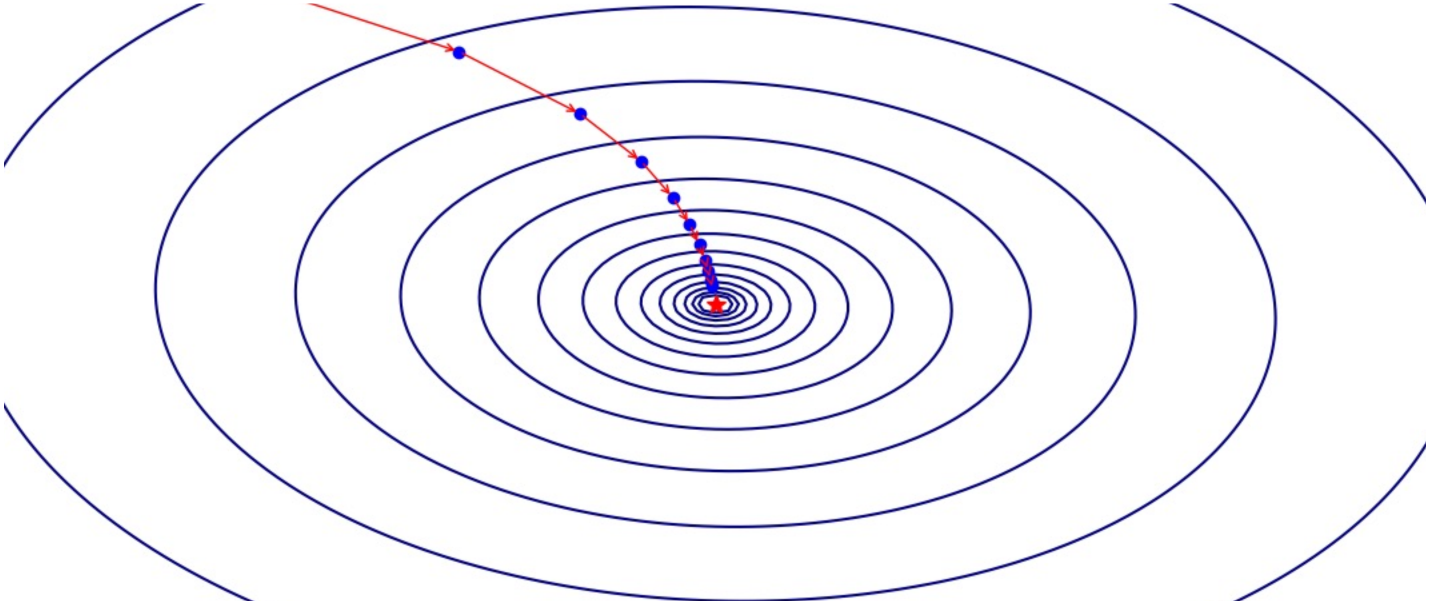


Machine Learning in Signal Processing

Winter Semester 2023/24

1. Introduction

17.10.2023

Prof. Dr. Vasileios Belagiannis

Chair of Multimedia Communications and Signal Processing

Friedrich-Alexander-Universität
Technische Fakultät

Today's Agenda

- Course Topics.
- Course Organisation.
- Team.
- Introduction to machine learning.

Course Topics

Not for sharing (LMS, Friedrich-Alexander-Universität Erlangen-Nürnberg)

1. **Introduction.**
2. Basics and terminology.
3. Linear regression.
4. Linear classification.
5. Performance evaluation.
6. Neural networks.
7. Deep neural networks.
8. Decision trees.
9. Ensemble models.
10. Random forests.
11. Clustering / Unsupervised learning.
12. Dimensionality reduction.
13. Support vector machines.
14. Recap and Q&A.

- The exam will be written.
- We will have an exam preparation test before the end of the year.

Acknowledgements

Ideas and inspiration from:

- CSC311 Introduction to Machine Learning, University of Toronto.
- Introduction to Machine Learning: LMU Munich.
- Introduction to Machine Learning, CSAIL, MIT.
- CSE 574 Introduction to Machine Learning, University of Buffalo.
- Special thanks Arij Bouazizi, Julia Hornauer, Julian Wiederer, Adrian Holzbock and Youssef Dawoud for contributing to the lecture preparation.

Course Exercises

- Machine learning by hand.
- Classification problems.
- Regression problems.
- Neural networks.
- Random forest.
- Support vector machines.

Overview

- This course introduces machine learning where background on machine learning is not excepted.
- We will learn specific machine learning algorithms and techniques for training and evaluating an approach.
- Background: linear algebra, calculus and programming.
- 2 + 2 SWS (90 minutes lecture + 90 minutes exercises).
- Schedule: Lecture on Tuesdays at 08:15 and exercise on Thursdays at 16:15.
- Material online at http://studon.fau.de
- The slides will be uploaded in advance.
- Communication: email / studon.

Team

- Lectures: Vasileios Belagiannis.
- Exercises: Amir El-Ghoussani.
- Special thanks to Arij Bouazizi, Adrian Holzbock, Julia Hornauer, Julian Wiederer, Youssef Dawoud for the slide creation support.

About the instructor

Studies

– Democritus University of Thrace (Dipl. Eng.), 2004 – 2009
– Technische Universität München (M.Sc.), 2009 – 2011
– Technische Universität München (Dr. rer. nat.), 2011 – 2015

Experience

– VGG, University of Oxford (Post-doc), 2015 – 2017
– Vision, OSRAM (Senior Researcher / Post-doc), 2017– 2018
– Universität Ulm (Assistant Professor), 2018 – 2022
– Universität Magdeburg (Professor), 2022
– FAU (Professor), 2022 -

Machine Learning @ FAU

– Chair of Multimedia Communications and Signal Processing.

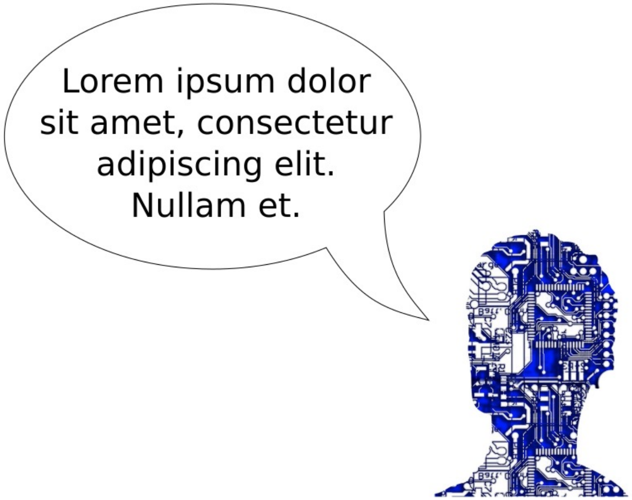
Literature

- Hastie, Trevor, et al. The elements of statistical learning: data mining, inference, and prediction. Vol. 2. New York: springer, 2009.
- Shalev-Shwartz, Shai, and Shai Ben-David. Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.
- Raschka, Sebastian, et al. Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python. Packt Publishing Ltd, 2022.

What is machine learning?

Not for sharing (LMS, Friedrich-Alexander-Universität Erlangen-Nürnberg)

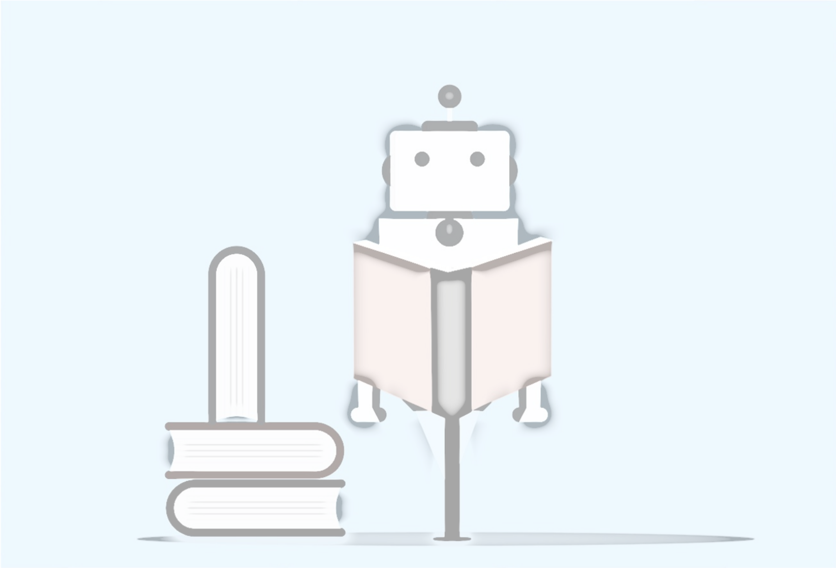
- A computer program which performs actions for which it was not explicitly programmed.
- Instead, it was given data and an objective function to learn the performed actions.
- Arthur Samuel: “Field of study that gives computers the ability to learn without being explicitly programmed”*.



*Samuel, Arthur L. "Some studies in machine learning using the game of checkers. II—Recent progress." IBM Journal of research and development 11.6 (1967): 601-617.

Why Learning from Data?

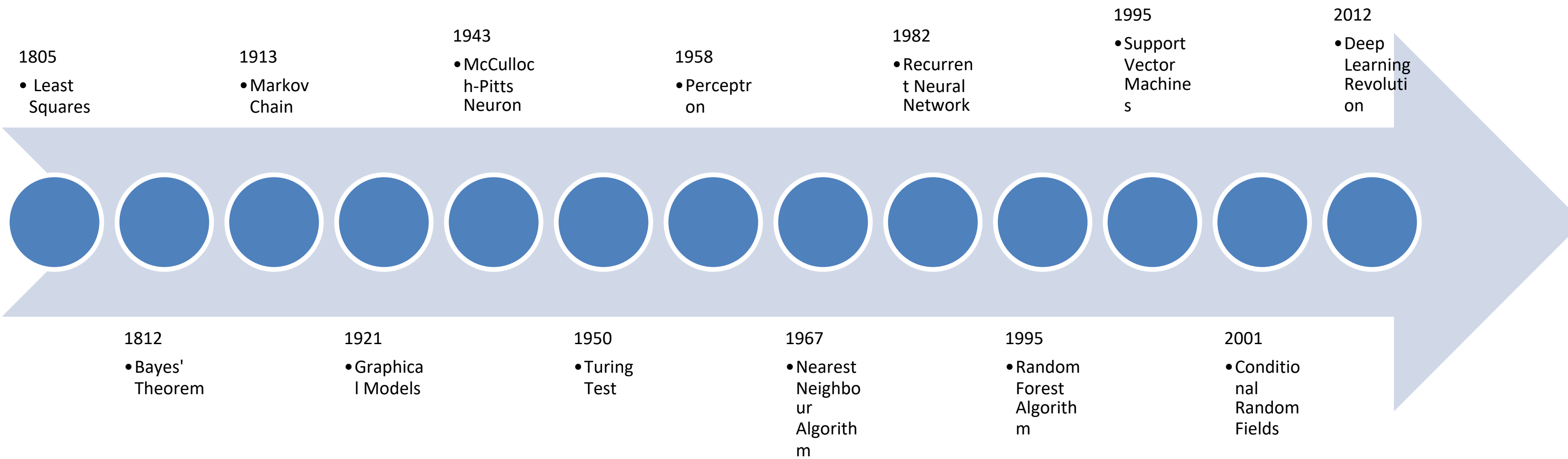
- Analytical solution might not exist.
- Closed-form solution might be too expensive.
- Machine learning mostly builds on numerical methods to define optimisations for different types of supervision.
- *What does the supervision term mean in the context of machine learning?*



By mohamed_hassan https://pixabay.com/de/vectors/machine-learning-bücher-algorithmus-6079971/

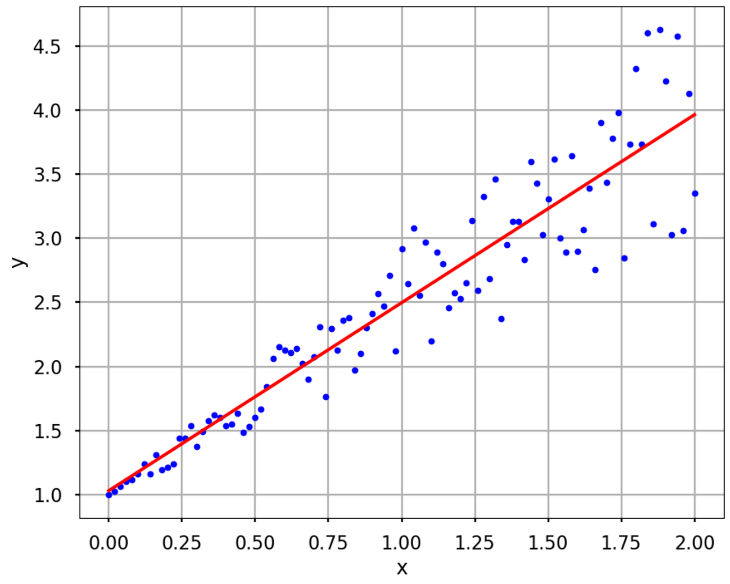
Online reference: https://machinelearningmastery.com/analytical-vs-numerical-solutions-in-machine-learning/

Machine Learning Milestones (Timeline)



Least Squares (Adrien-Marie Legendre, 1805)

- Find the best-fitting curve for a set of points, e.g. (x, y).
- *How?*
  - Minimizing the sum of the squares of the point offsets (residuals) from the the curve $f(\cdot)$.
  - Linear fit curve: $f(a, b) = a + bx$.
  - n points.
  - Residual minimisation $S = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$.
  - Set the gradient of S to zero and solve for each unknown term.
  - *How many unknown terms?*

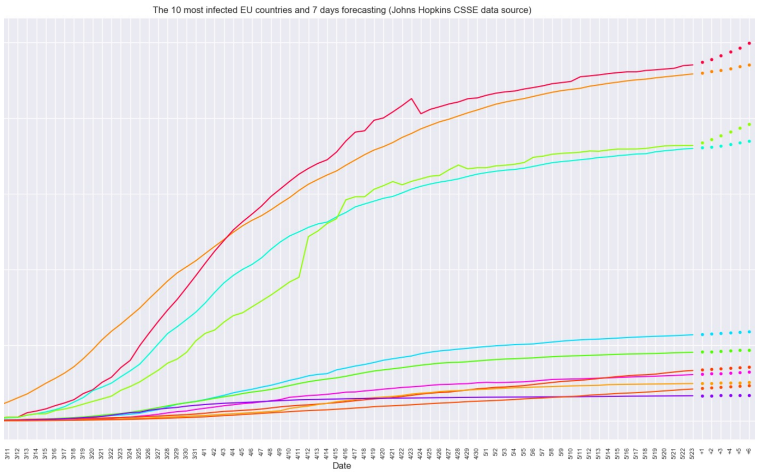


Online reference: https://mathworld.wolfram.com/LeastSquaresFitting.html

Least Squares (Cont.)

Not for sharing (LMS, Friedrich-Alexander-Universität Erlangen-Nürnberg)

- Carl Friedrich Gauss studied the problem too (1795).
- Linear and non-linear least squares.
  - Depends on whether the residuals are linear.
- Linear least squares have <u>closed-form solution</u>.
- Least square are used for <u>regression</u> problems.
- *Can we predict the COVID-19 cases with a linear regression model?*



Source: https://github.com/bazilas/COVID19-EU

Python Examples:
https://python4mpia.github.io/fitting_data/least-squares-fitting.html
https://pythonnumericalmethods.berkeley.edu/notebooks/chapter16.04-Least-Squares-Regression-in-Python.html

Bayes' Theorem (1812)

- A principled way to calculate the conditional probability $P(A|B)$ without requiring the joint probability $P(A, B)$.
- $P(A|B)$ is known as posterior probability.
- $Posterior = \frac{Likelihood * Prior}{Evidence}$.
- Named after the statistician and philosopher Thomas Bayes.
- It is often used in machine learning for the maximum a posteriori estimation (MAP).

Prior probability: Probability of A.

Likelihood: Probability of B being true, while A is true.

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

Evidence: Probability of B.

Online reference: https://machinelearningmastery.com/bayes-theorem-for-machine-learning/

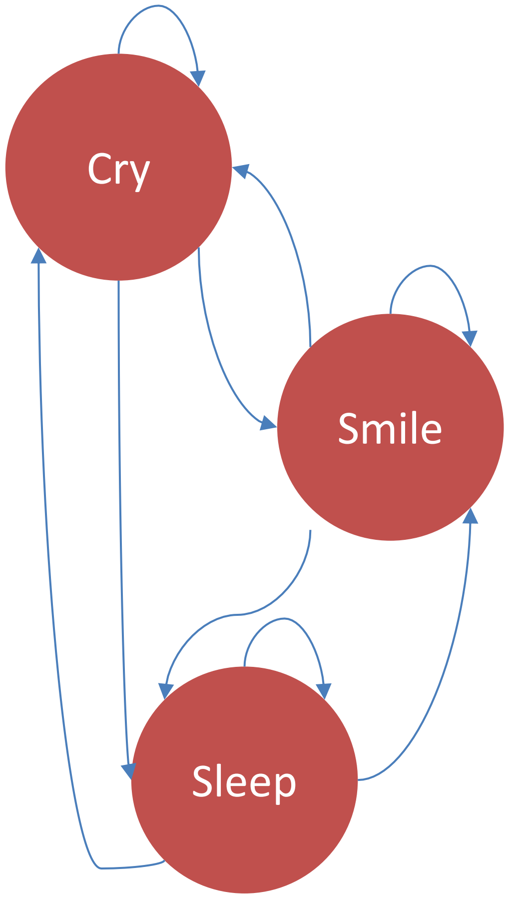
Bayes' Theorem (Cont.)

- What is the probability that a senior citizen to die in 2020? (fictional example)
  - About 2.5 million of 280 million citizens died during 2020.
  - About 1.5 out of 17 million senior citizens died during 2020.
  - Unconditional probability:
    - $P(A) = \frac{2.5}{280} = 0.0089$ (population mortality rate).
  - Conditional probability:
    - $P(B) = \frac{17}{280} = 0.0607$ (probability to be senior citizen).
    - $P(A, B) = \frac{1.5}{280} = 0.0053$ (probability of senior citizens who died).
    - $P(A|B) = \frac{P(A,B)}{P(B)} = 0.0873$ (probability of a senior citizen to die – among senior citizens).
    - $P(B|A) = \frac{P(A,B)}{P(B)} = 0.5955$ (probability of a senior citizen to die in total population).
  - Bayes' Theorem:
    - $P(A|B) = \frac{P(B|A)P(A)}{P(B)} = 0.0873$ (probability of a senior citizen to die).

Example Motivation: https://plato.stanford.edu/entries/bayes-theorem/

Markov Chain (1913)

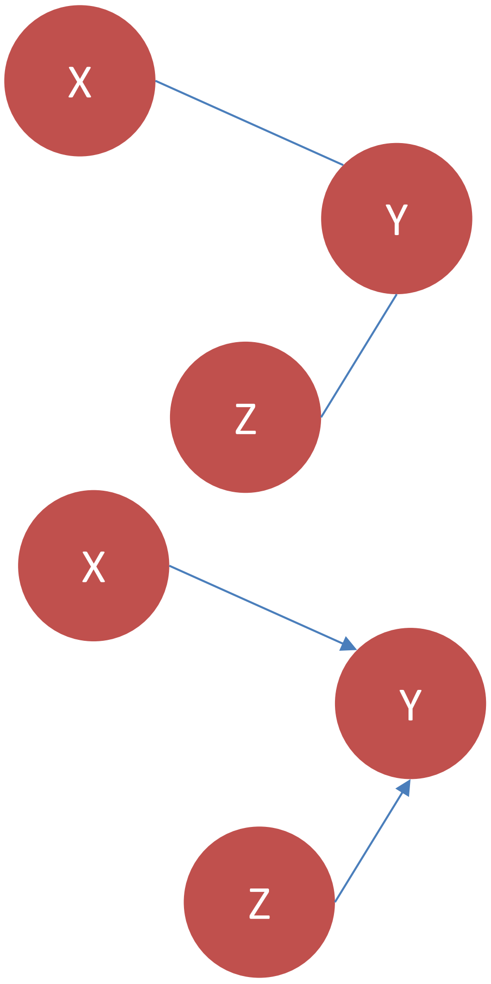
- A Markov chains is a stochastic/random process that is memory-less.
  - A system of states with transition probabilities between the states.
  - Compute the next step based on the current one.
  - Fixed state space, e.g. cry, smile and sleep. Goal predict what will happen next.
  - Applications: finance, genetics, computer vision, machine learning.
- Named after the mathematician Andrey Andreyevich Markov.



Online Demo: https://setosa.io/ev/markov-chains/

Graphical Models (1921)

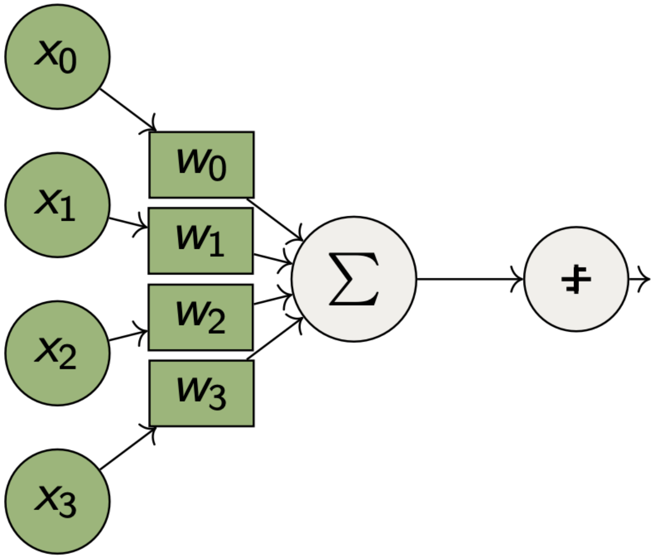
- A. A. Markov, Josiah Willard Gibbs and Sewall Wright independently come up with graphical model formulations.
- They are also called probabilistic graphical model (PGM).
- The graph expresses the conditional dependencies between random variables, e.g. X, Y, Z.
- Undirected graphical models and Bayesian networks are common graphical models in the literature.



Online reference: http://www.math.chalmers.se/~wermuth/pdfs/encybehav.pdf

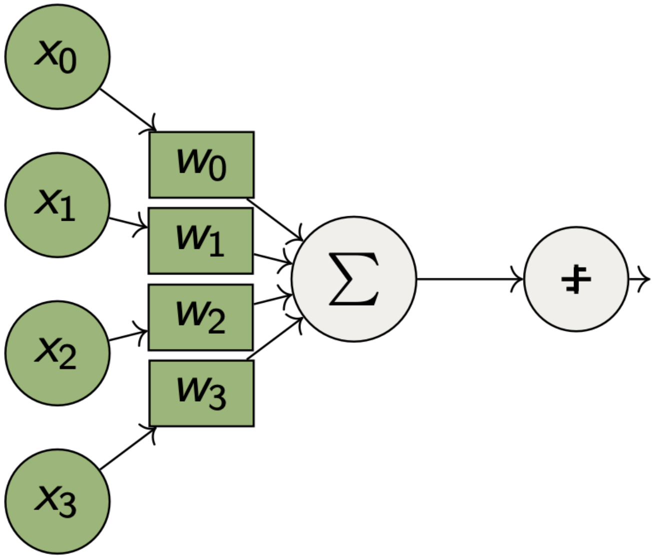
McCulloch-Pitts Neuron (1943)

- Neurophysiologist Warren McCulloch & logician Walter Pitts.
- Input $x_i \in \{0,1\}$ and output $y \in \{0,1\}$.
- Excitatory or inhibitory input, represented by the weights/parameters $w_i \in \{-1,1\}$.
- The mapping is represented by a threshold function $f: \mathbb{R}^D \to \mathbb{R}$.
  - $f(x) = \begin{Bmatrix} 0, if\ \boldsymbol{w}x \leq T \\ 1, otherwise \end{Bmatrix}.$



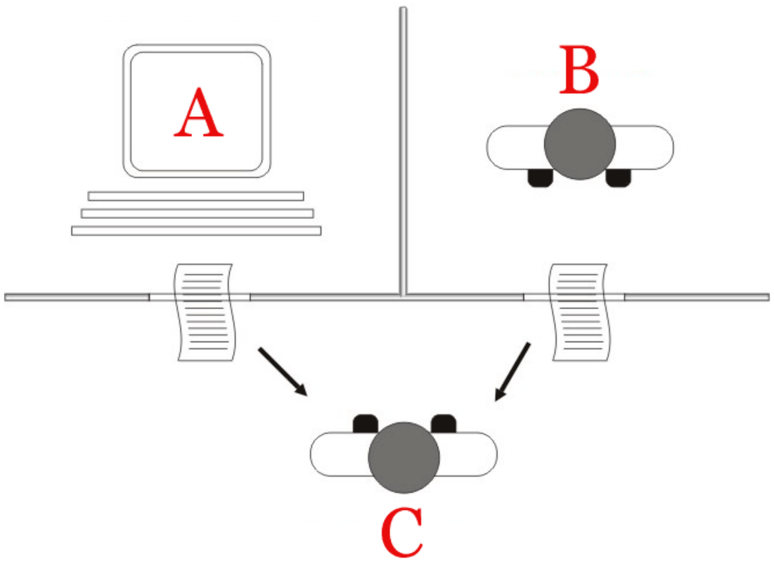
McCulloch-Pitts Neuron (Cont.)

- A few observations.
  - Weighted sum as output.
  - No learning algorithm.
  - The weights and threshold are hand-computed in advance.
- Limitations:
  - Only 0-1 input.
  - Hand-defined parameters.
  - XOR or complex gates cannot be computed.
  - Known as Threshold Logic Unit (first artificial unit).



Turing Test and (1950)

- Proposed by Alan Mathison Turing*.
  - Logician and computer pioneer with major work in artificial intelligence.
  - Invented the Turing machine (1936), an abstract machine that is able to implement any computer algorithm.
- Test question: “Can machines think?”
- To answer the question, the imitation game was proposed by Alan Turing.

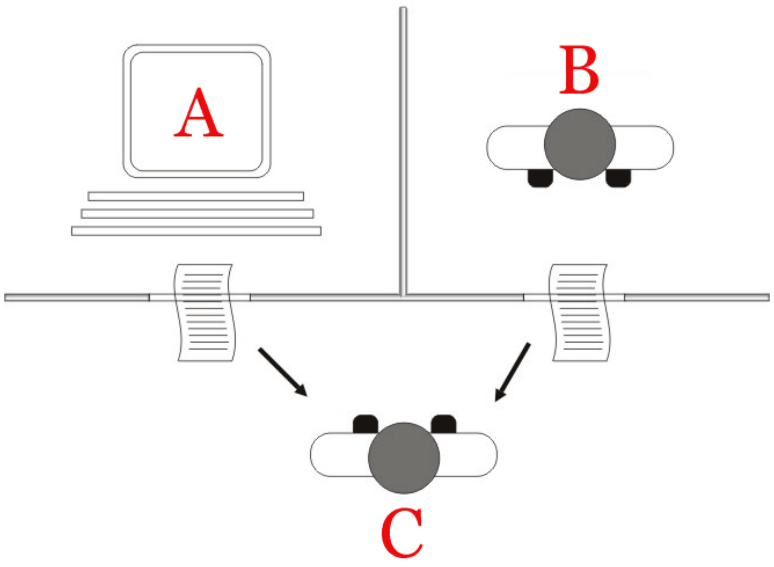


By Juan Alberto Sánchez Margallo - https://commons.wikimedia.org/wiki/File:Test_de_Turing.jpg, CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=57298943

*Turing, Alan M. "Computing machinery and intelligence." *Parsing the Turing test*. Springer, Dordrecht, 2009. 23-65.

Turing Test and (Cont.)

- Originally called: imitation game.
- Player A: computer.
- Player B: human.
- Player C: interrogator.
- Goal:
  - C tries to understand whether A or B is a computer.
- Rules:
  - A,B and C are separated from each other.
  - Message-based communication only.

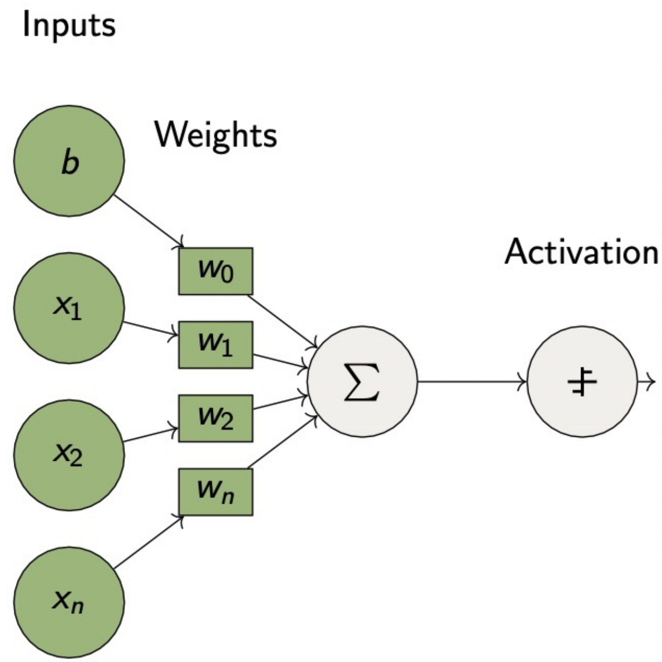


By Juan Alberto Sánchez Margallo - https://commons.wikimedia.org/wiki/File:Test_de_Turing.jpg, CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=57298943

*Turing, Alan M. "Computing machinery and intelligence." *Parsing the Turing test*. Springer, Dordrecht, 2009. 23-65.

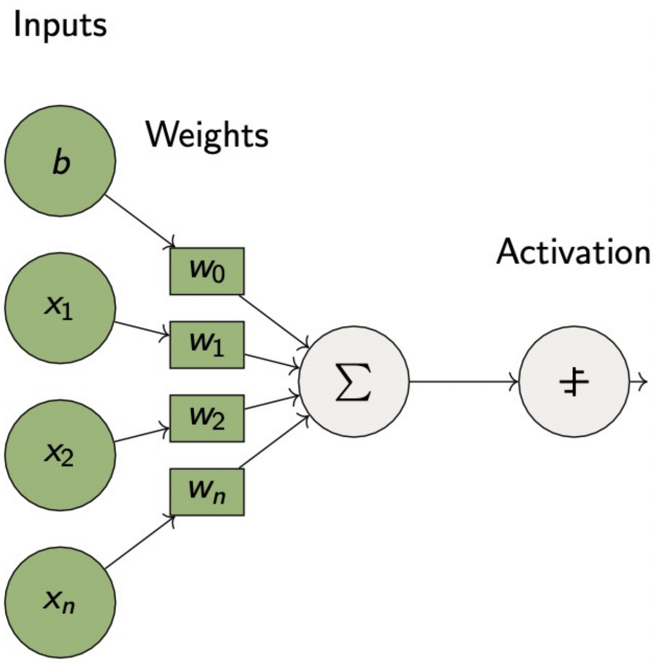
Perceptron (1958)

- Developed by Frank Rosenblatt (Psychologist)
- A linter classifier for binary problems.
- Real-valued input vector.
- Supervised learning based on n inputs and output pairs.
  - $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$.
- Represented by a threshold function as:
  - $f(x) = \begin{Bmatrix} 1, if\ \boldsymbol{w}x + b \geq 0 \\ 0, \quad otherwise \end{Bmatrix}$.
- The parameters of the Perceptron are learned.



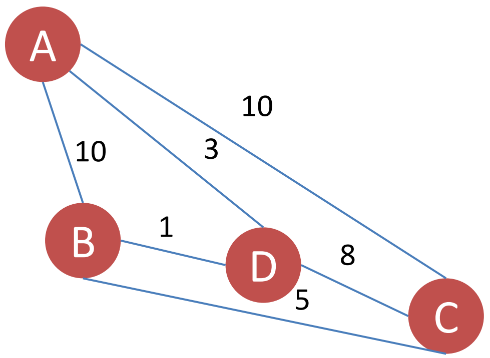
Perceptron (Cont.)

- Learning rule:
  - The parameter $\boldsymbol{w}$ update is proportional to the input and the difference between the prediction $f(x_i)$ and ground-truth y_i .
- Algorithm for single-layer Perceptron:
  1. Initialize the parameters $\boldsymbol{w}$ and set learning rate η and threshold.
  2. Update the weights: $w_j(t+1) = w_j(t) + \eta\big(y_i - f(x_i)\big)x$.
  3. Iterate until convergence.
- Implemented at Mark I Perceptron machine.
- Multiclass Perceptron was later developed.



Nearest Neighbour Algorithm (1967)

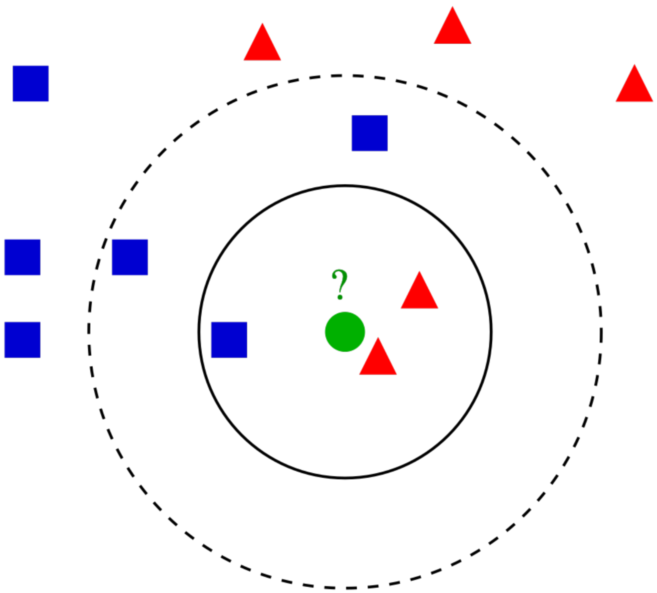
- One of the first algorithms to solve the travelling salesman problem with an approximation*.
- Applied to graph models in general.
- It gives a greedy solution that is not necessarily the optimal.
- Algorithm:
  1. Mark all vertices as unvisited and pick randomly one to visit.
  2. Find the next unvisited vertex with the smallest cost and visit.
  3. Repeat 2 until all vertices haven been visited.



Online example: https://www.people.vcu.edu/~gasmerom/MAT131/nearest.html

Nearest Neighbour Algorithm (Cont.)

- k-NN: k-Nearest Neighbour algorithm.
  - Given a reference sample, find the k nearest neighbours.
- Application on classification, e.g. the green point will be classified as triangle if we consider the 3 closest nearest neighbours.
- Application to regression problems.
- *How do we apply the algorithm for regression problems?*

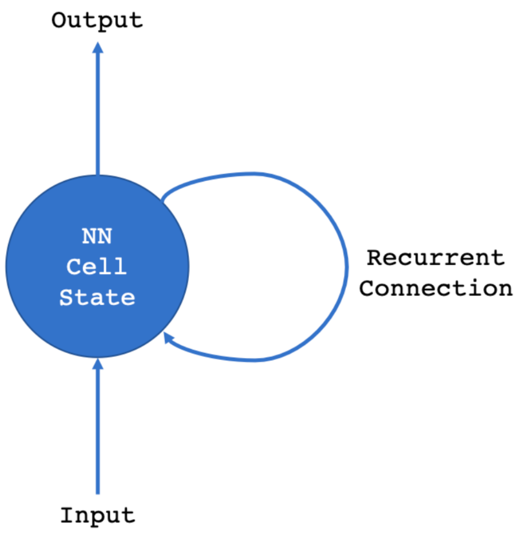


By Antti Ajanki AnAj - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=2170282

*Cover, Thomas, and Peter Hart. "Nearest neighbor pattern classification." *IEEE transactions on information theory* 13.1 (1967): 21-27.

Recurrent Neural Network (1982)

- Artificial neural networks with dynamic behaviour.
- A special type of Hopfield network.
- Thanks to their memory, they process long input sequences.
- In theory, they are Turing complete (simulate any Turing machine).
- An RNN can be unrolled and represented as a feed-forward neural network.
- Long short-term memory (LSTM) networks are the most popular variant of RNNs.

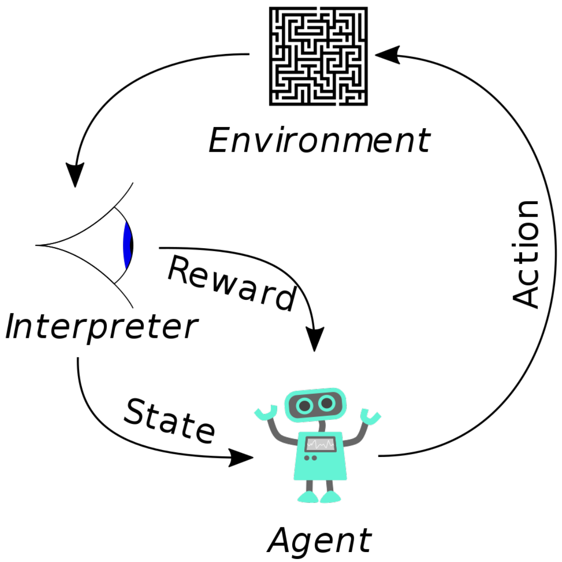


Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

Reinforcement Learning (1989)

Not for sharing (LMS, Friedrich-Alexander-Universität Erlangen-Nürnberg)

- An agent perform a sequence of actions to maximize the cumulative reward.
- Learn how to map a situation to actions*.
- The agent does not know the actions in advance.
- Exploration and exploitation of actions is necessary.
  - Exploit known actions to maximize the reward.
  - Explore new actions to further improve the reward.

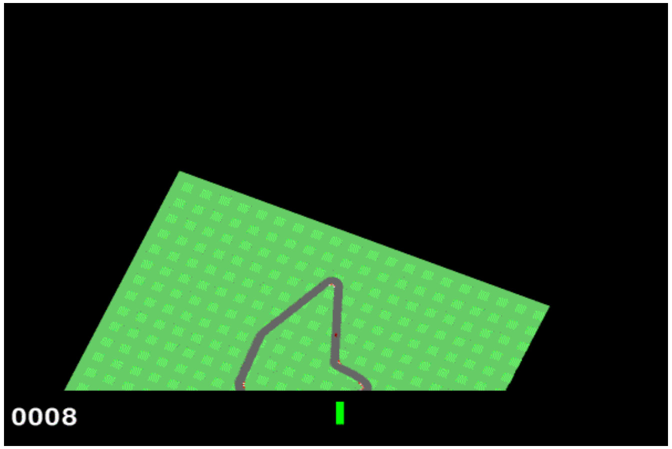


By Megajuice - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=5789574

*Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 2018 (Chapter 1).

Reinforcement Learning (1989)

- Standard reinforcement learning is modelled as a Markov decision process:
  - Agent.
  - Environment.
  - Set of actions.
  - Transition probability from one state to another one given an action.
  - Immediate reward after transition from one state to another one given an action.
- Goal: learn a policy to maximize the expected cumulative reward.



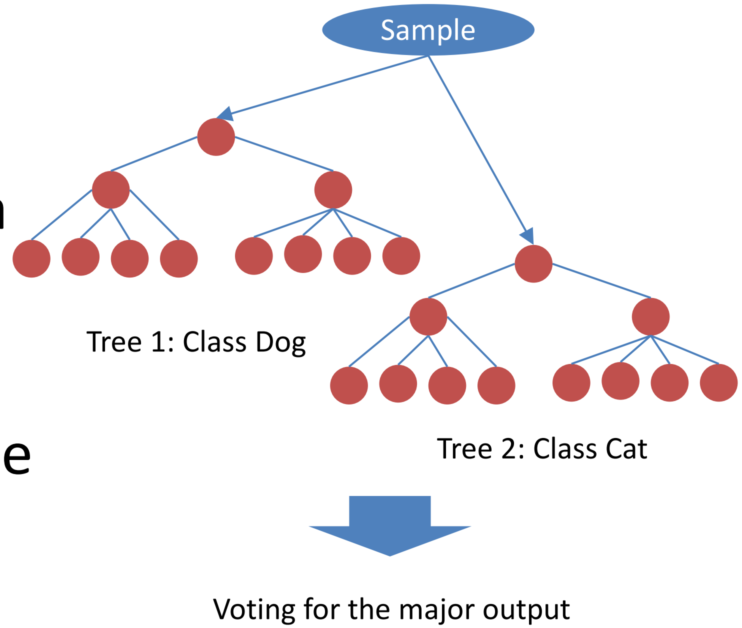
By Oleg Klimov, https://www.gymlibrary.dev/environments/box2d/car_racing/#credits

*Sutton, Richard S. "Learning to predict by the methods of temporal differences." Machine learning 3.1 (1988): 9-44. APA.

Random Forest Algorithm (1995)

Not for sharing (LMS, Friedrich-Alexander-Universität Erlangen-Nürnberg)

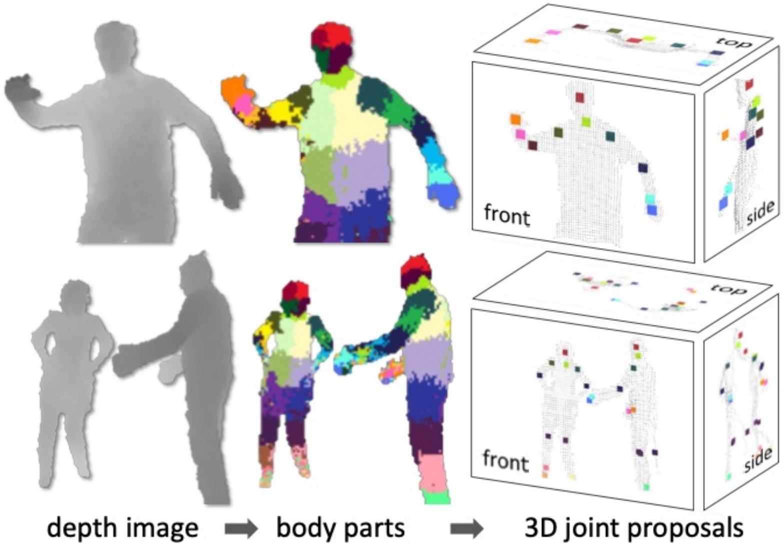
- Ensemble method for classification and regression.
- Multiple trees are learned with different subsets of the training data for each one.
- Tin Kam Ho proposed the first algorithm*, but the approach became popular from Leo Breiman's** extension to the algorithm (2001).
- Applied on classification and regression.



*Ho, Tin Kam. "Random decision forests." Proceedings of 3rd international conference on document analysis and recognition. Vol. 1. IEEE, 1995.
**Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.

Random Forest Algorithm (Cont.)

- Depth image input and random forest were used to regression the 3D body joints*.
  - Large-scale dataset for training the forest.
  - Application on MS Kinect©.
- *What is the forest output during the regression of body parts?*
- Random forest has been also successfully used in finance, medical imaging, and e-commerce applications.

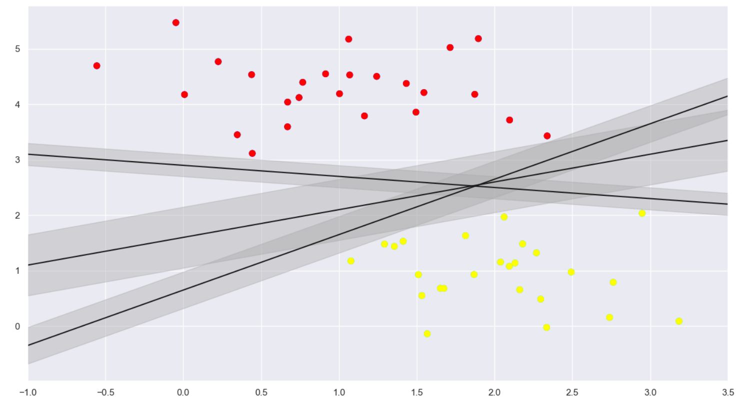


Source: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/BodyPartRecognition.pdf

*Shotton, Jamie, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. "Real-time human pose recognition in parts from single depth images." In CVPR 2011, pp. 1297-1304. Ieee, 2011.

Support Vector Machines (1995)

- Supervised learning algorithms for classification and regression.
- Developed by Corinna Cortes and Vladimir Vapnik.
- In classification, the goal is to separate the data with lines or planes.
- Some of the training samples compose the support vector to perform the separation.
- Linear case: lines to separate.
- Non-linear cases: kernels to separate.

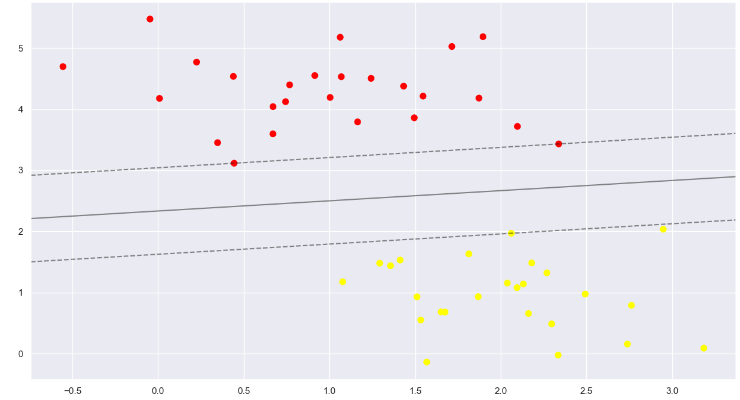


2-class Problem (linear).

Online example: https://jakevdp.github.io/PythonDataScienceHandbook/05.07-support-vector-machines.html

Support Vector Machines (1995)

- Support vector machines became the standard tool for classification in computer vision in the past.
  - Object detection was performed with HOG features and SVM classifiers*.
  - Now replaced by Neural Networks.
- SVMs are still used tough for many other applications, e.g. stock market prediction.
- SVMs were also combined with Conditional random fields (2001) for structured prediction**.



2-class Problem (linear).

*Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." (CVPR'05). Vol. 1. Ieee, 2005.
**Lafferty, John, Andrew McCallum, and Fernando CN Pereira. "Conditional random fields: Probabilistic models for segmenting and laabeling sequence data." (2001).

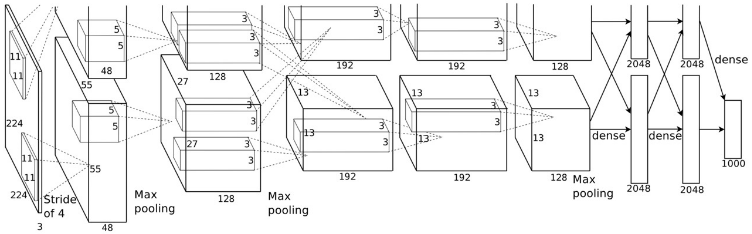
Deep Learning Revolution (2012)

Not for sharing (LMS, Friedrich-Alexander-Universität Erlangen-Nürnberg)

- Deep learning is a family of methods to learn data representations, inference and sampling models.
  - Possibilities: features, embeddings, classifiers, regressors, in general predictors and samplers among others.
- Parametric models with:
  - A composition of linear and non-linear functions.
  - One or multiple objectives to learn the parameters from data.
- Represented by deep neural networks.

Deep Learning Revolution (2012)

- AlexNet* significantly contributed to the revolution of deep learning.
- It has been the winner of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012.
- Since then much deeper neural networks have appeared in the literature.

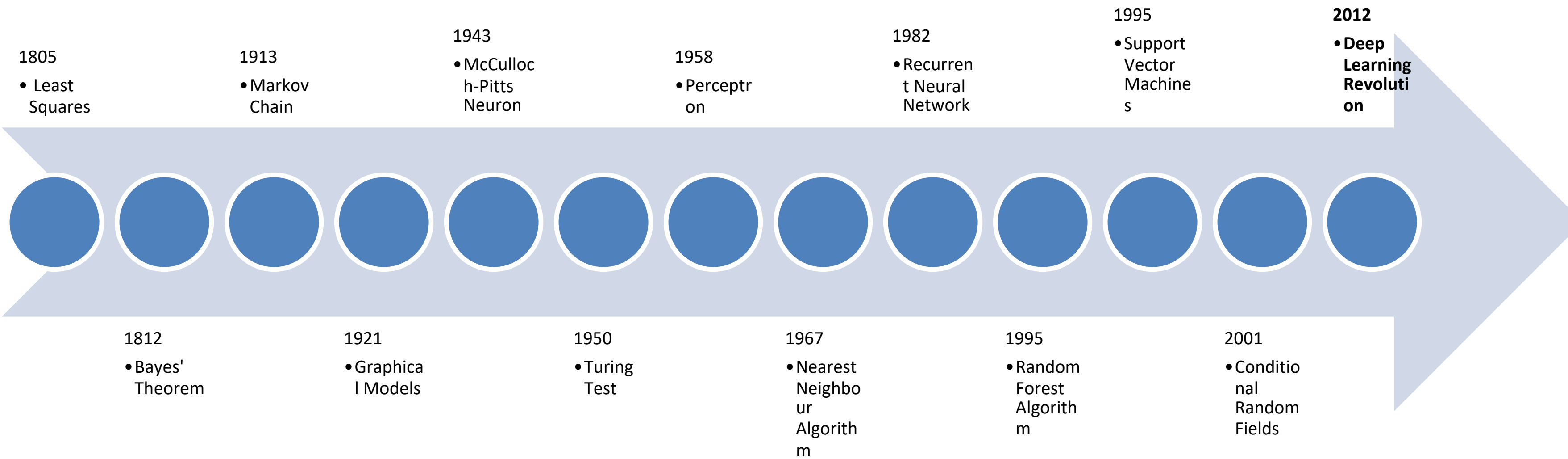


Source: https://www.semanticscholar.org/paper/ImageNet-classification-with-deep-convolutional-Krizhevsky-Sutskever/abd1c342495432171beb7ca8fd9551ef13cbd0ff

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60.6 (2017): 84-90.

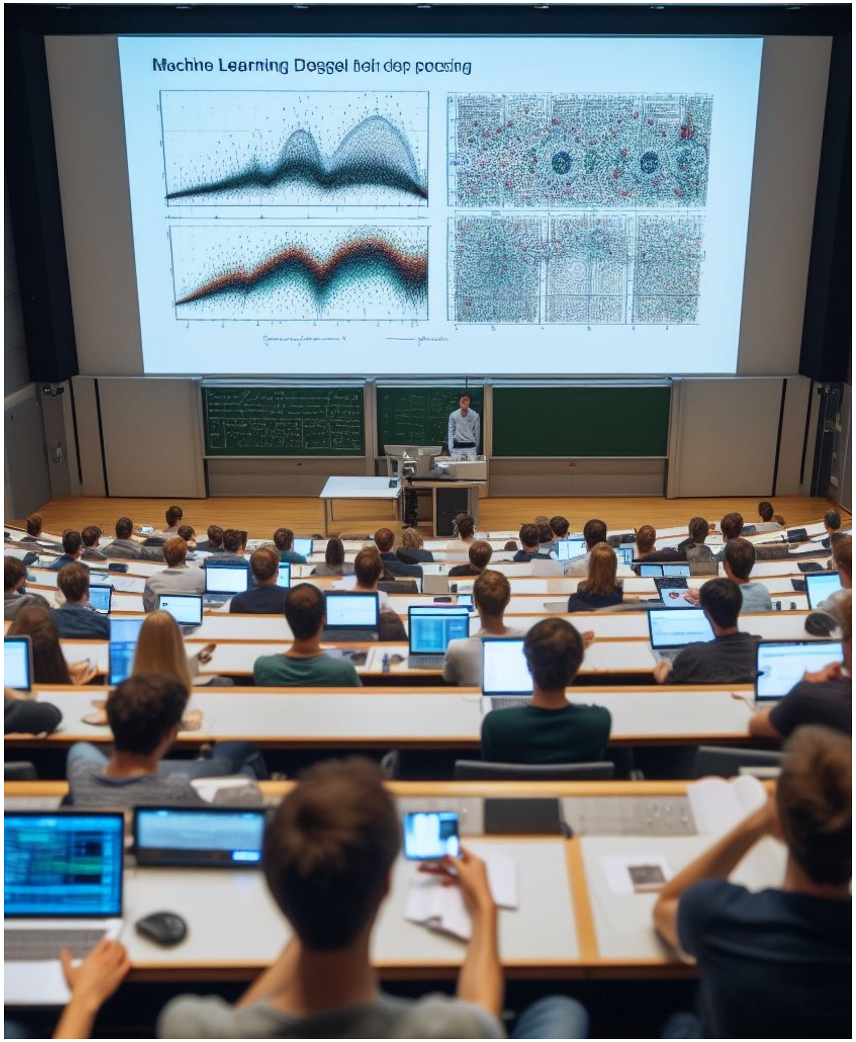
Machine Learning Milestones (Timeline)

- Our goal is to study the most important milestones in machine learning.



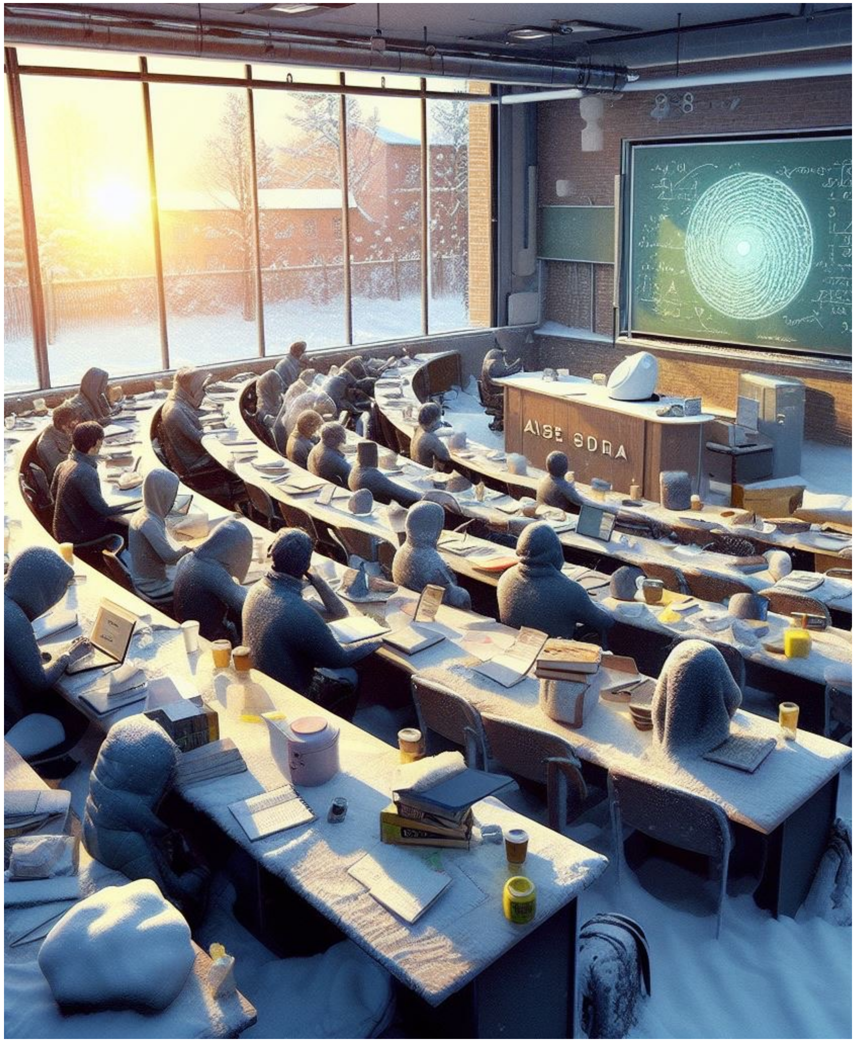
Machine Learning Today

- Generated with Bing Image Creator (DALLE 3)
  - https://www.bing.com/images
- Input prompt:
  - *Lecture on Machine Learning in Signal Processing at the University of Erlangen.*



Machine Learning Today (Cont.)

- Generated with Bing Image Creator (DALLE 3)
  - https://www.bing.com/images
- Input prompt:
  - *First day at the Machine Learning lecture early in the morning on a cold day.*



Next Lecture

Basics & Terminology.