Linear Algebra

Paul Rad, Ph.D.

Associate Professor
Cyber Analytics and AI
Information Systems and Cyber Security
College of Business School
210.872.7259

A Few Quotes

“A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)

“Machine learning is the next Internet”
(Tony Tether, Director, DARPA)

Machine learning is the hot new thing”
(John Hennessy, President, Stanford)

“Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)

A Short History of Machine Learning

1950 — **Alan Turing** creates the “Turing Test” to determine if a computer has real intelligence. To pass the test, a computer must be able to fool a human into believing it is also human.

1952 — **Arthur Samuel** wrote the first computer learning program. The program was the game of checkers, and the IBM computer improved at the same the more it played studying which moves made up winning and incorporating those moves into its program.

1957 — **Frank Rosenblatt** designed the first neural network for computers (the perceptron), which simulate the thought processes of the human brain.

1967 — The “nearest neighbor” algorithm was written, allowing computers to begin using very basic pattern recognition. This could be used to map a route for traveling salesmen, starting at a random city but ensuring they visit all cities during a short tour.

1979 — Students at Stanford University invent the “Stanford Cart” which can navigate obstacles in a room on its own.

1981 — **Gerald Dejong** introduces the concept of Explanation Based Learning (EBL), in which a computer analyses training data and creates a general rule it can follow by discarding unimportant data.

1985 — **Terry Sejnowski** invents NetTalk, which learns to pronounce words the same way a baby does.

A Short History of Machine Learning

1990s — Work on machine learning shifts from a knowledge-driven approach to a data-driven approach. Scientists begin creating programs for computers to analyze large amounts of data and draw conclusions — or “learn” — from the results.

1997 — IBM’s Deep Blue beats the world champion at chess.

2006 — Geoffrey Hinton coins the term “deep learning” to explain new algorithms that let computers “see” and distinguish objects and text in images and videos.

2010 — The Microsoft Kinect can track 20 human feature at a rate of 30 times per second, allowing people to interact with the computer via movements and gestures.

2011 — IBM’s Watson beats its human competitors at Jeopardy.

2011 — Google Brain is developed, and its deep neural network can learn to discover and categorize objects much the way a cat does.

2012 – Google’s X Lab develops a machine learning algorithm that is able to autonomously browse YouTube videos to identify the videos that contain cats.

A Short History of Machine Learning

2014 – Facebook develops DeepFace, a software algorithm that is able to recognize or verify individuals on photos to the same level as humans can.

2015 – Amazon launches its own machine learning platform.

2015 – Microsoft creates the Distributed Machine Learning Toolkit, which enables the efficient distribution of machine learning problems across multiple computers.

2015 – Over 3,000 AI and Robotics researchers, endorsed by Stephen Hawking, Elon Musk and Steve Wozniak (among many others), sign on open letter warning of the danger of autonomous weapons which select and engage targets without human intervention.

2016 – Google's artificial intelligence algorithm beats a professional player at the Chinese board game Go, which is considered the world's most complex board game and is many times harder than chess. The AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Go competition.

Machine Learning Definition

Artificial Intelligence (**AI**) is intelligence exhibited by machines. An ideal "intelligent" machine is a flexible rational agent that perceives its environment and takes actions **that maximize its chance of success at some goal**.

In 1959, Arthur Samuel **defined machine learning** as a "Field of study that gives computers the ability to learn without being **explicitly programmed**". **Machine learning** explores the study and construction of algorithms that can learn from and make predictions on data.

Machine learning is a subfield of computer science that evolved from the study of **pattern recognition** and **computational learning theory** in artificial intelligence

Machine learning addresses the question of how to build computers that improve automatically through experience. It is one of today's most rapidly growing technical fields, lying at the intersection of computer engineer/science and statistics, and at the core of artificial intelligence and data science.

Recent progress in machine learning has been driven both by the development of new learning algorithms and by the ongoing explosion in the availability of data and low-cost computation (cloud computing).

The adoption of data-intensive machine-learning methods can be found throughout science, technology and commerce, leading to more evidence-based decision-making across many walks of life, including health care, manufacturing, education, financial modeling, and marketing.

Machine Learning Definition

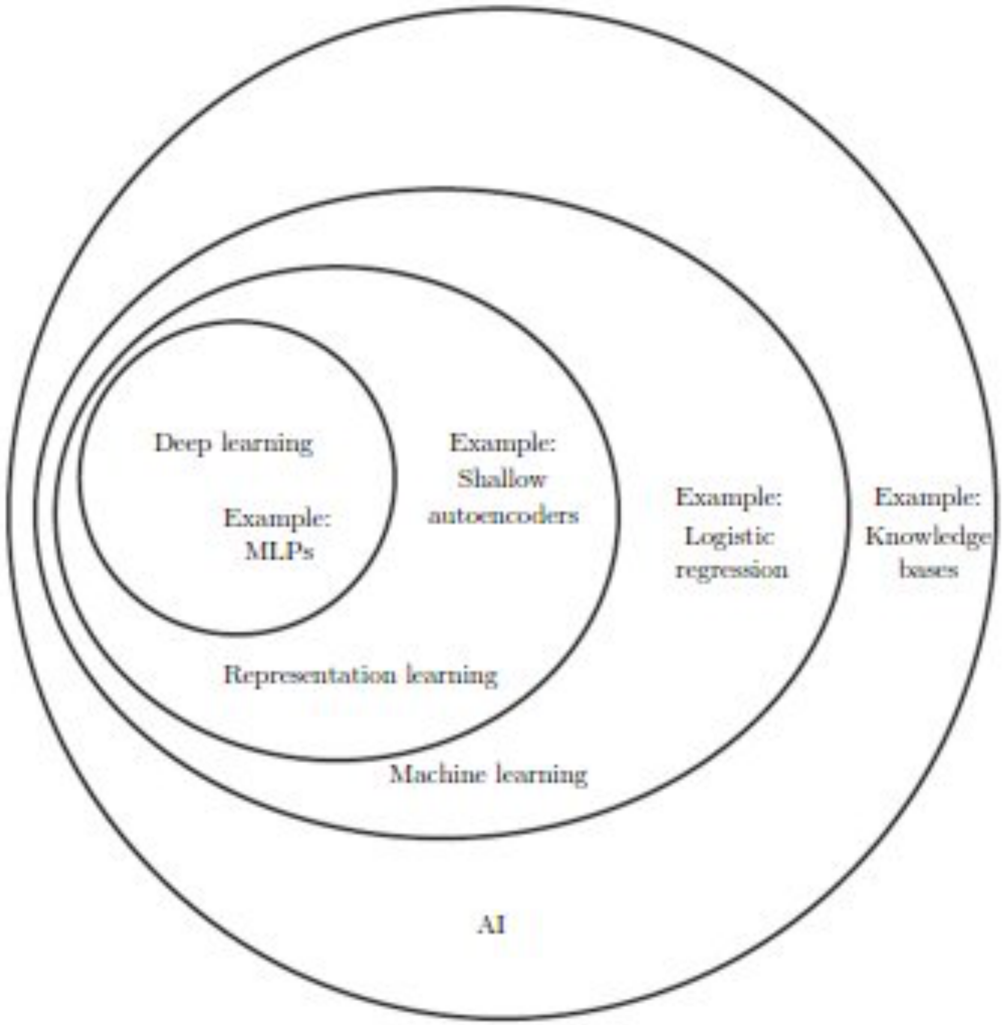
A machine learning algorithm is an algorithm that is able to learn from data.

Tom Mitchell (1997) provides the definition:

"A computer program is said to learn from

experience *E* with respect to some class of task *T* and performance measure *P*,

if its performance at tasks in *T*, as measured by *P*, improves with experience *E*."



Component of learning

Credit approval

Applicant information form

Age	Attributes	23 Years
Gender	x1	Male
Annual Salary	x2	$30,000
Years in residence	x3	1 year
Years in job	x4	1 year
Current debt	x5	$15,000
….		….

Approve credit?

Learning System Model (Inductive Learning)

Given examples of a function *(X, F(X))*

Predict function *F(X)* for new examples *X*

Discrete *F(X)*: Classification

Continuous *F(X)*: Regression

F(X) = Probability(*X*): Probability estimation

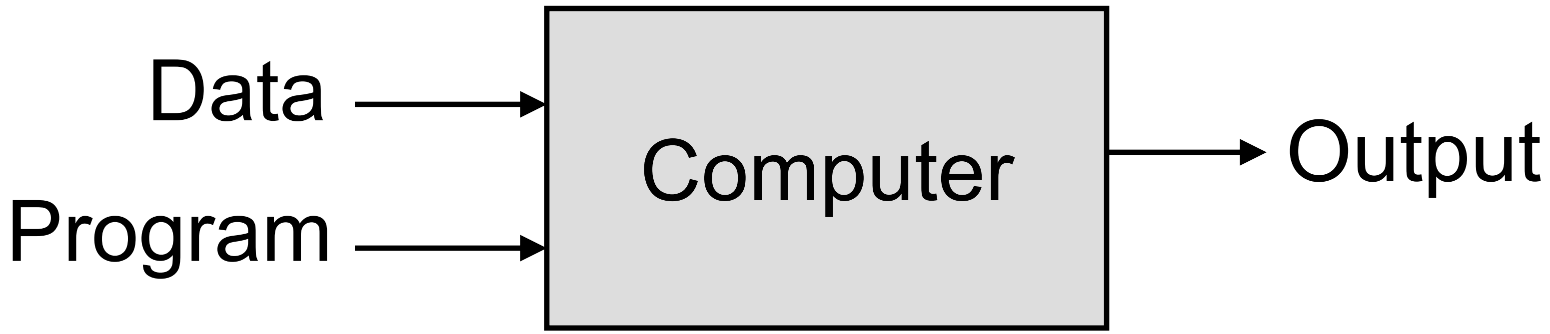
Component of Learning

- Input: $\mathbf{x}$ Customer Application
- Output: y Good/bad customer
- Target function: $f : \mathcal{X} \to \mathcal{Y}$ Ideal credit approval formula
- Data: $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)$ History records

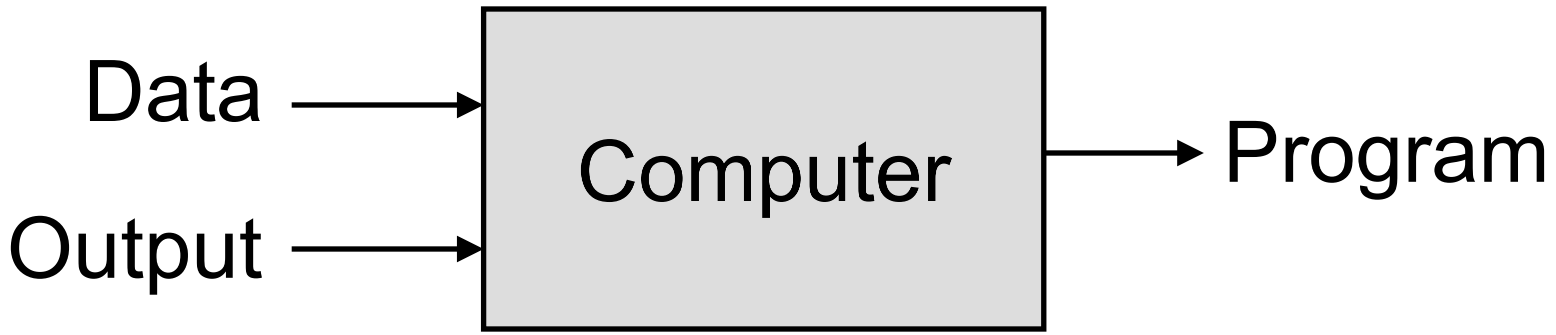
$$\downarrow \quad \downarrow \quad \downarrow$$

- Hypothesis: $g : \mathcal{X} \to \mathcal{Y}$ Formula to be used

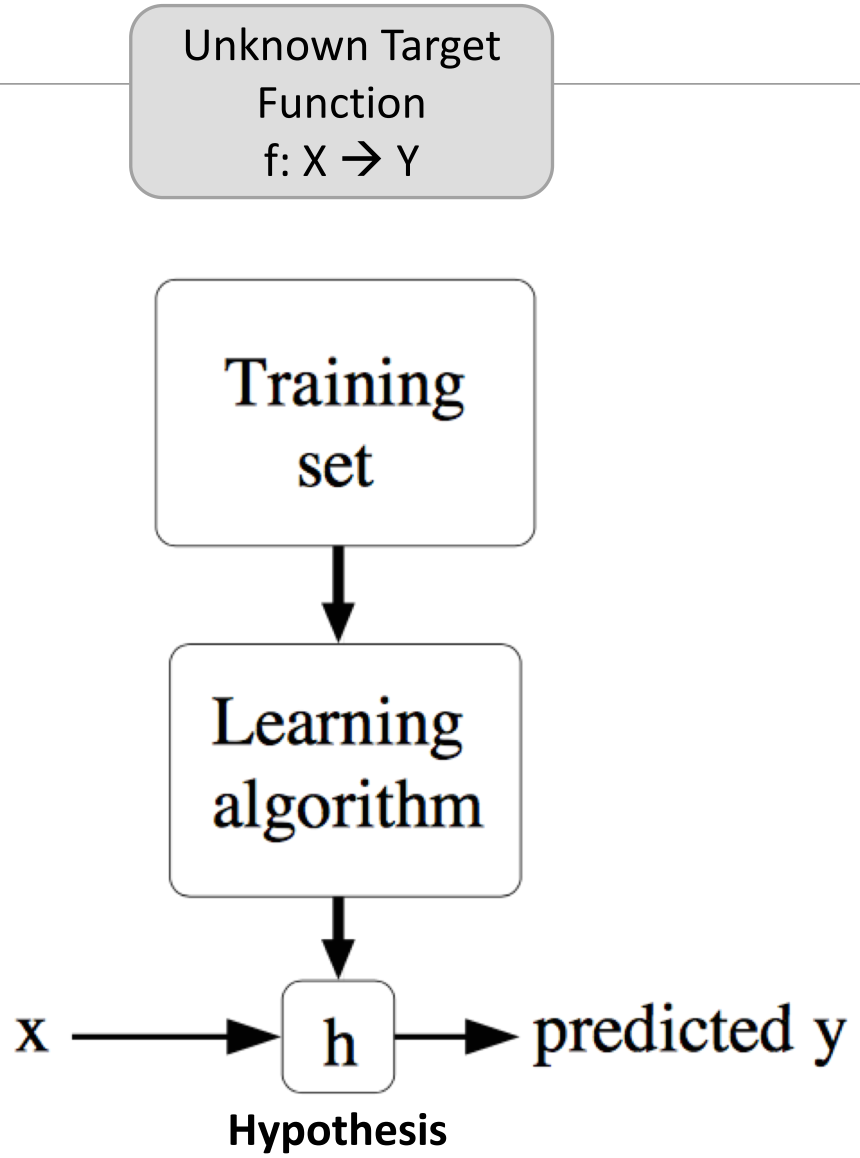
Traditional Programming



Machine Learning



Learning Model



A related experiment

- Consider a 'bin' with red and green marbles.

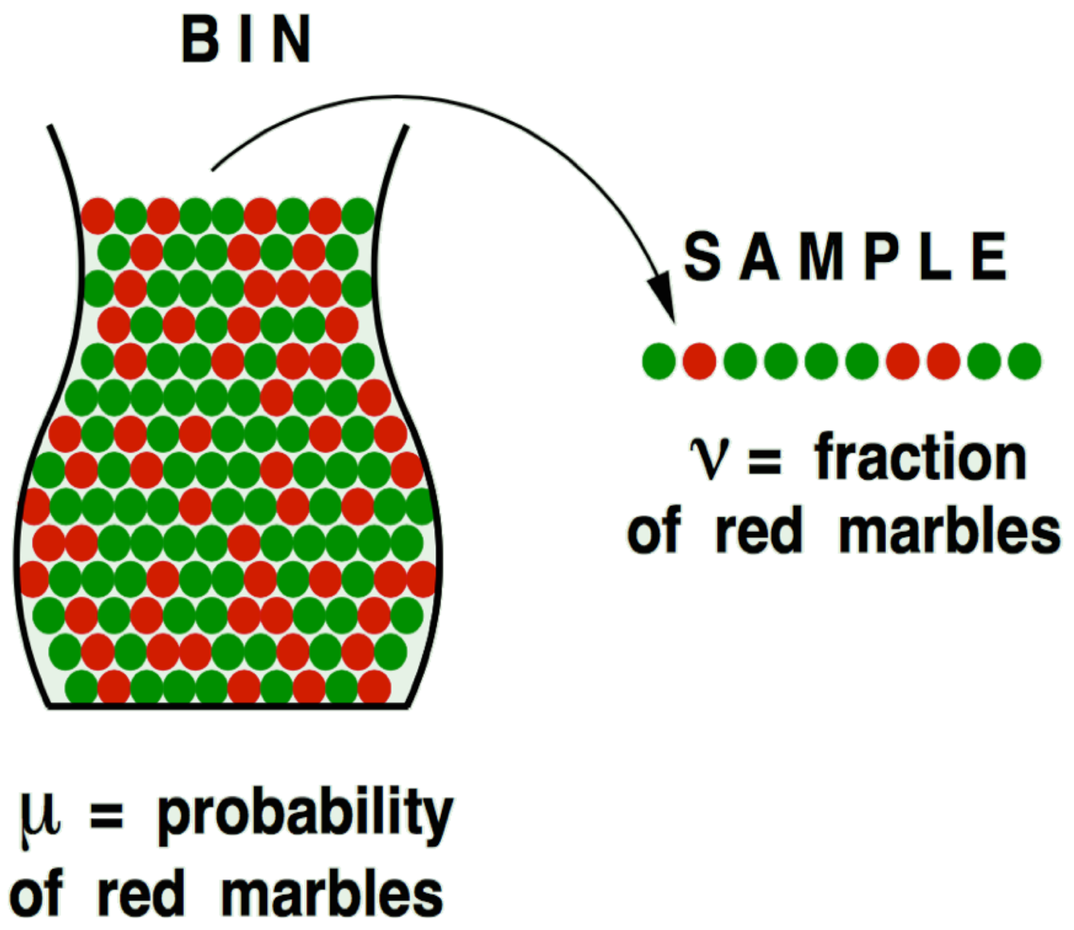
$$\mathbb{P}[\text{ picking a red marble }] = \mu$$

$$\mathbb{P}[\text{ picking a green marble }] = 1 - \mu$$

- The value of μ is unknown to us.

- We pick N marbles independently.

- The fraction of red marbles in sample $= \nu$



What <u>does</u> ν say about μ?

In a big sample (large N), ν is probably close to μ (within ϵ).

Formally,

$$\mathbb{P}\left[\,|\nu - \mu| > \epsilon\,\right] \leq 2e^{-2\epsilon^2 N}$$

This is called Hoeffding's Inequality.

A Simple Algorithm " the perceptron"

For input $\mathbf{x} = (x_1, \cdots, x_d)$ 'attributes of a customer'

$$\text{Approve credit if} \quad \sum_{i=1}^{d} w_i x_i > \text{threshold},$$

$$\text{Deny credit if} \quad \sum_{i=1}^{d} w_i x_i < \text{threshold}.$$

This linear formula $h \in \mathcal{H}$ can be written as

$$h(\mathbf{x}) = \text{sign}\left(\left(\sum_{i=1}^{d} w_i x_i\right) - \text{threshold}\right)$$

A Simple Algorithm " the perceptron"

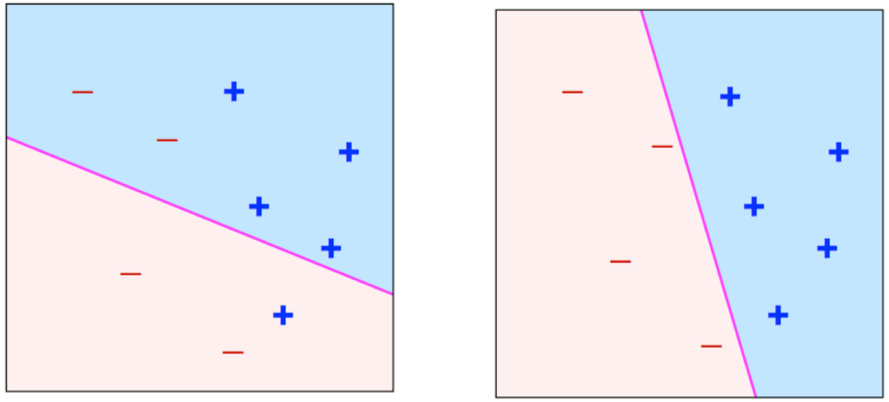
$$h(\mathbf{x}) = \text{sign}\left(\left(\sum_{i=1}^{d} w_i \; x_i\right) + \quad w_0 \quad\right)$$

Introduce an artificial coordinate $x_0 = 1$:

$$h(\mathbf{x}) = \text{sign}\left(\sum_{i=0}^{d} w_i \; x_i\right)$$

In vector form, the perceptron implements

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^{\intercal}\mathbf{x})$$



'linearly separable' data

A Simple Algorithm " the perceptron"

The perceptron implements

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^\intercal \mathbf{x})$$

Given the training set:

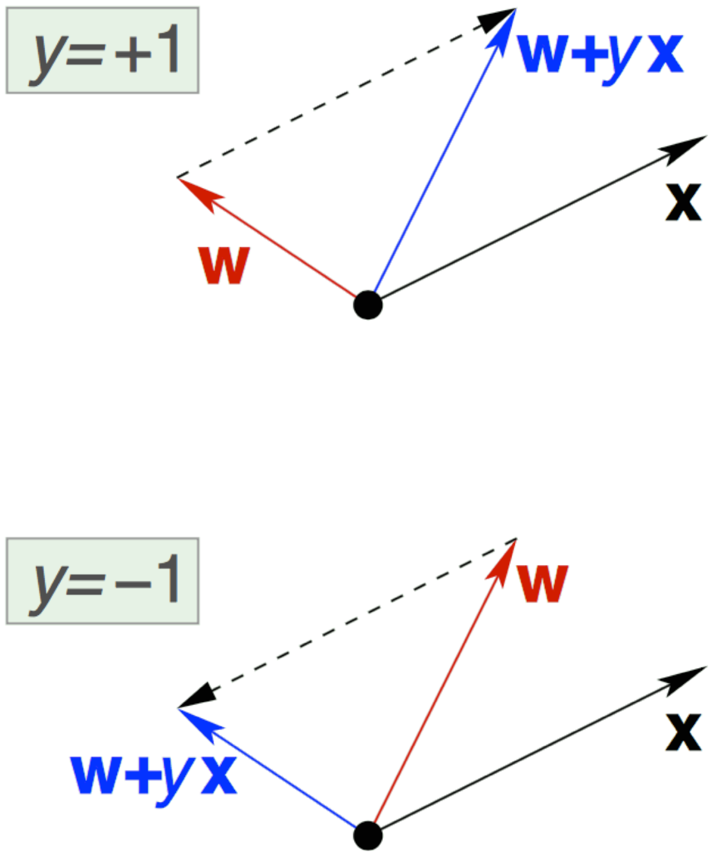
$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)$$

pick a misclassified point:

$$\text{sign}(\mathbf{w}^\intercal \mathbf{x}_n) \neq y_n$$

and update the weight vector:

$$\boxed{\mathbf{w} \leftarrow \mathbf{w} + y_n \mathbf{x}_n}$$



PLA Iteration

- One iteration of the PLA:

$$\mathbf{w} \leftarrow \mathbf{w} + y\mathbf{x}$$

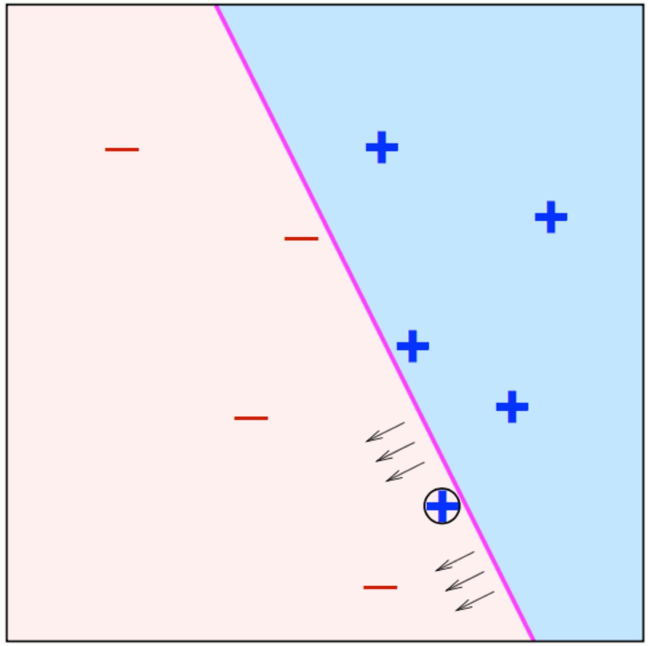
where $(\mathbf{x}, y)$ is a misclassified training point.

- At iteration $t = 1, 2, 3, \cdots$, pick a misclassified point from

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)$$

and run a PLA iteration on it.

- That's it!



Algorithm

The success of machine learning system 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.

Thank you