Machine Learning (ML)
and
Deep Learning

Hyuk Lim

Energy AI Track
Korea Institute of Energy Technology

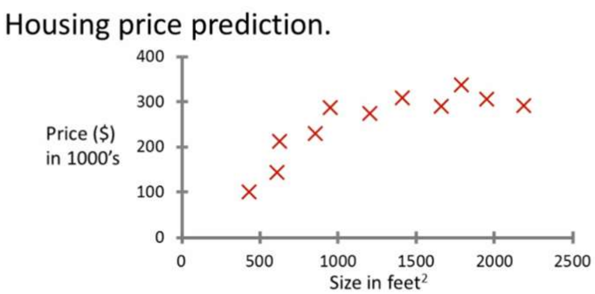
Contents

- Introduction
- Artificial Neural Network (ANN)
- Deep Learning
- Machine Learning Algorithms

Introduction

Problems – Regression

- Finding the relationship between one dependent variable and a series of other variables (independent variables)

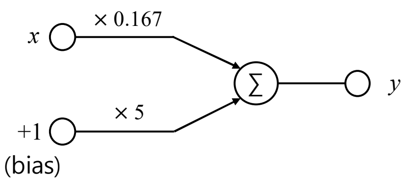


https://www.holehouse.org/mlclass/01_02_Introduction_regression_analysis_and_gr.html

Problems – Regression

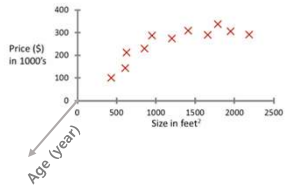
- Linear regression

Graphical representation

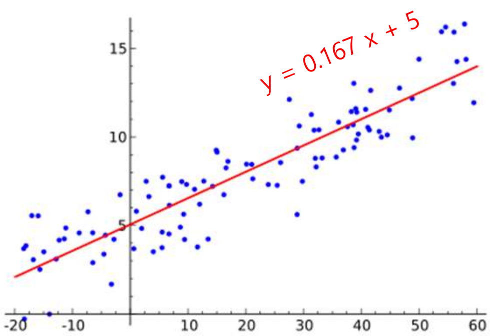


https://en.wikipedia.org/wiki/Linear_regression

Problems – Regression



- Linear regression



How to obtain the linear regression model?

$\hat{y} = \theta_0 + \theta_1 \cdot x_1$ 1-dimensional space

$\hat{y} = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2$ 2-dimensional space

$\vdots$

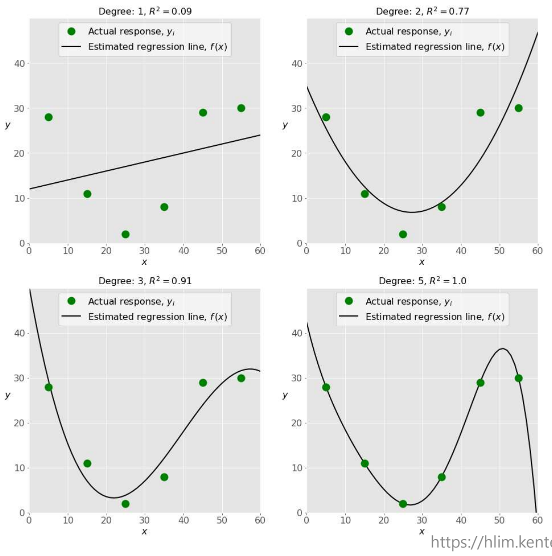
$\hat{y} = \theta_0 + \theta_1 \cdot x_1 + \cdots + \theta_n \cdot x_n$ n-dimensional space

Solution:

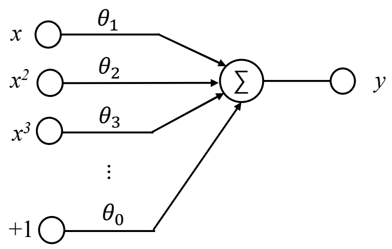
$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbf{y}$$

https://en.wikipedia.org/wiki/Linear_regression

Problems – Regression



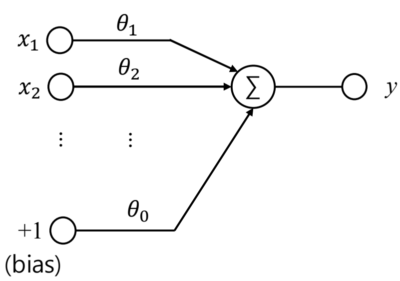
Graphical representation



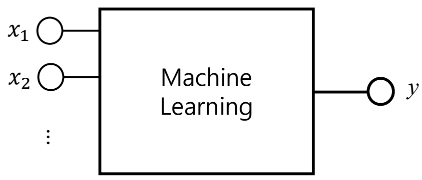
https://realpython.com/linear-regression-in-python/

Problems – Regression

- Linear regression vs machine learning

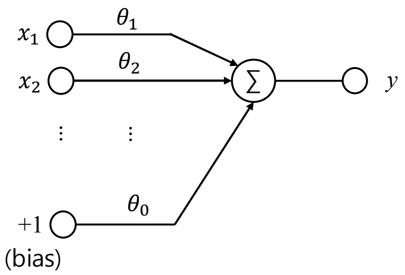


How to find weights?

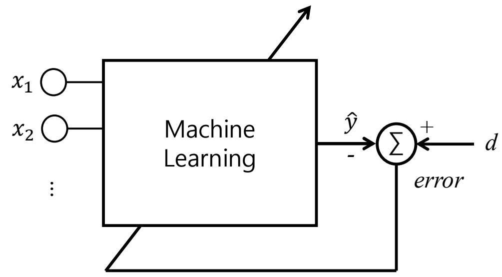


Problems – Regression

- Linear regression vs machine learning

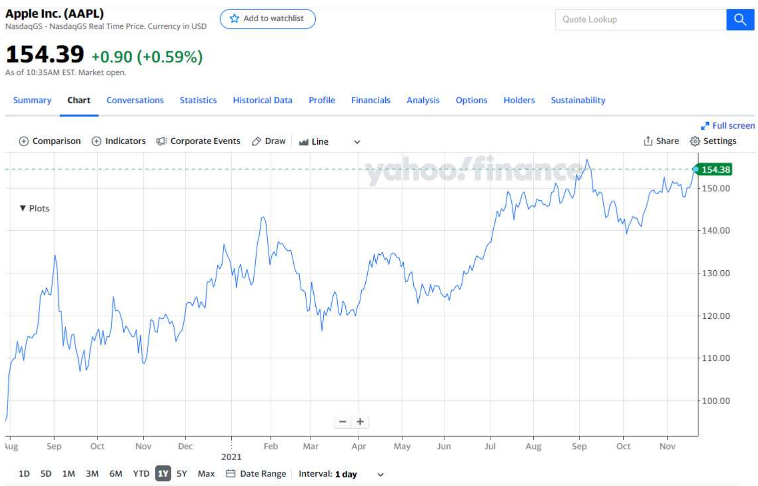


Update weights iteratively using the errors



Supervised learning

Problems – One/n-step Forecasting



What will be the stock price in tomorrow?

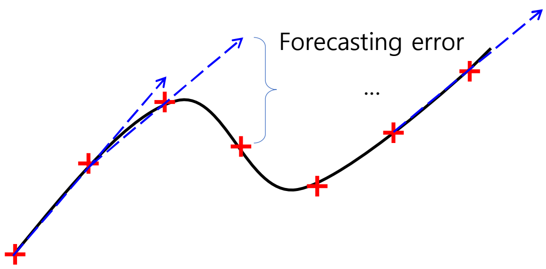
Problems – One/n-step Forecasting

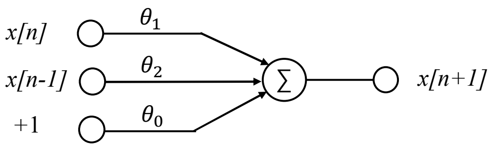
What will be the stock price in tomorrow?

Naïve solution:
Today price will be the tomorrow price.

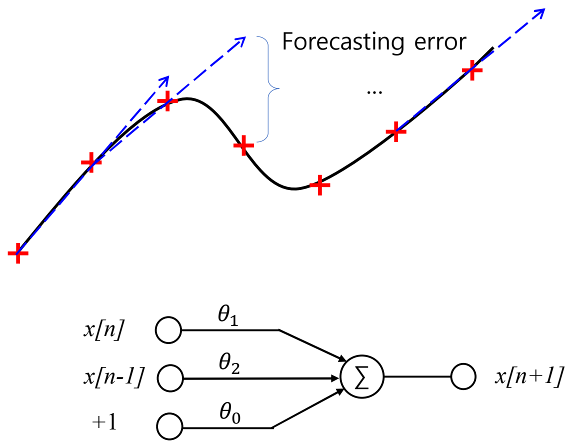
However, it could be higher or lower.

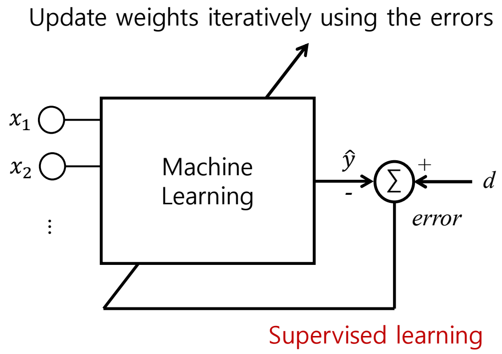
Problems – One/n-step Forecasting



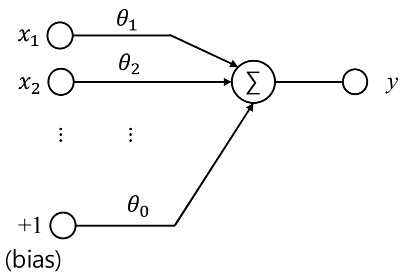


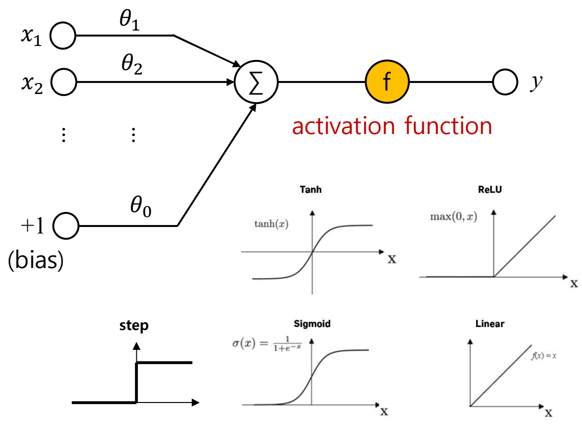
Problems – One/n-step Forecasting





Single Layer Perceptron

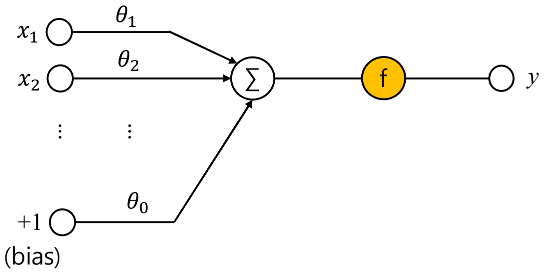




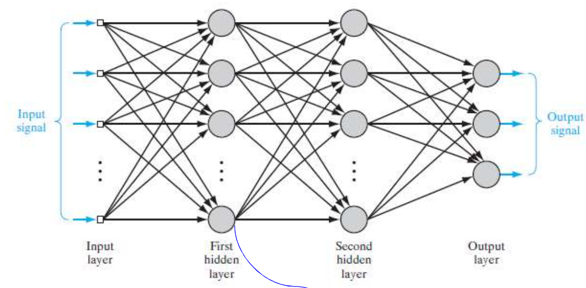
https://medium.com/@shrutijadon10104776/survey-on-activation-functions-for-deep-learning-9689331ba092

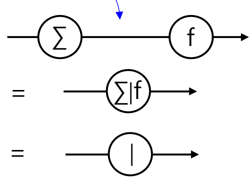
Multi Layer Perceptron (MLP)

Single layer perceptron



MLP

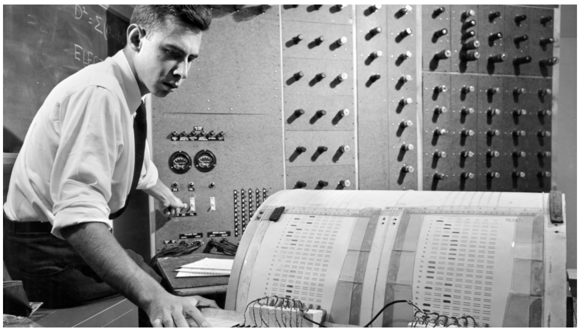




Artificial Neural Network (ANN)

Single Layer Perceptron

- Frank Rosenblatt (Ph.D. in Psychology, Cornell)



Vol. VI, No. 2, Summer 1958

1958

research trends

CORNELL AERONAUTICAL LABORATORY, INC., BUFFALO 21, NEW YORK

The Design of an Intelligent AUTOMATON

by FRANK ROSENBLATT

Introducing the perceptron — A machine which senses, recognizes, remembers, and responds like the human mind.

STORIES about the creation of machines having human qualities have long been a fascinating province in the realm of science fiction. Yet we are now about to witness the birth of such a machine — a machine capable of perceiving, recognizing, and identifying its surroundings without any human training or control.

Development of that machine has stemmed from a search for an understanding of the physical mechanisms which underlie human experience and intelligence. The question of the nature of these processes is at least as ancient as any other question in western science and philosophy, and, indeed, ranks as one of the greatest scientific challenges of our time.

Our understanding of this problem has gone perhaps as far as had the development of physics before Newton. We have some excellent descriptions of the phenomena to be explained, a number of interesting hypotheses, and a little detailed knowledge about events in the nervous system. But we lack agreement on any integrated set of principles by which the functioning of the nervous system can be understood.

We believe now that this ancient problem is about to yield to our theoretical investigation for three reasons:

First, in recent years our knowledge of the functioning of individual cells in the central nervous system has vastly increased.

Second, large numbers of engineers and mathematicians are, for the first time, undertaking serious study of the mathematical basis for thinking, perception, and the handling of information by the central nervous system, thus providing the hope that these problems may be within our intellectual grasp.

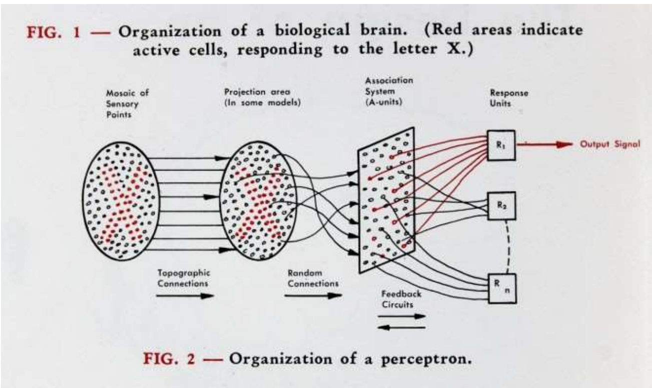
Third, recent developments in probability theory and in the mathematics of random processes provide new tools for the study of events in the nervous system, where only the gross statistical organization is known and the precise cell-by-cell "wiring diagram" may never be obtained.

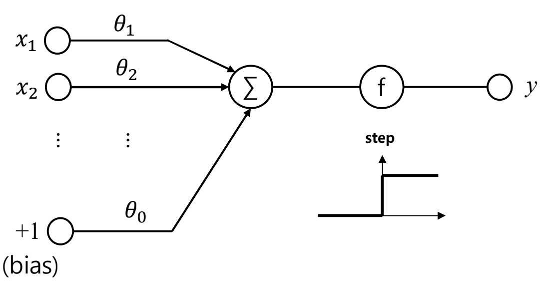
Receives Navy Support

In July, 1957, Project PARA (Perceiving and Recognizing Automaton), an internal research program which had been in progress for over a year at Cornell Aeronautical Laboratory, received the support of the Office of Naval Research. The program had been concerned primarily with the application of probability theory to

https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon

Single Layer Perceptron

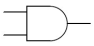




Rosenblatt's single-layer perceptron cannot classify input patterns that are not linearly separable.

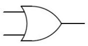
Logic Gates

- Logic gates: AND, OR, XOR, …



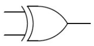
AND

A	B	Output
0	0	0
0	1	0
1	0	0
1	1	1



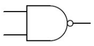
OR

A	B	Output
0	0	0
0	1	1
1	0	1
1	1	1



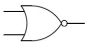
XOR

A	B	Output
0	0	0
0	1	1
1	0	1
1	1	0



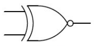
NAND

A	B	Output
0	0	1
0	1	1
1	0	1
1	1	0



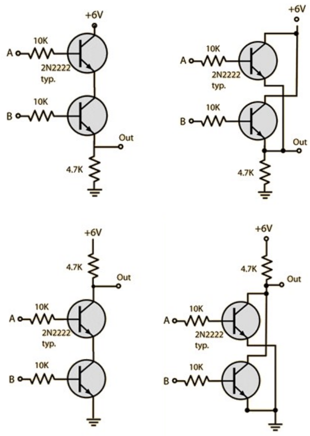
NOR

A	B	Output
0	0	1
0	1	0
1	0	0
1	1	0



XNOR

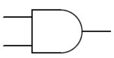
A	B	Output
0	0	1
0	1	0
1	0	0
1	1	1



https://instrumentationtools.com/logic-gates/

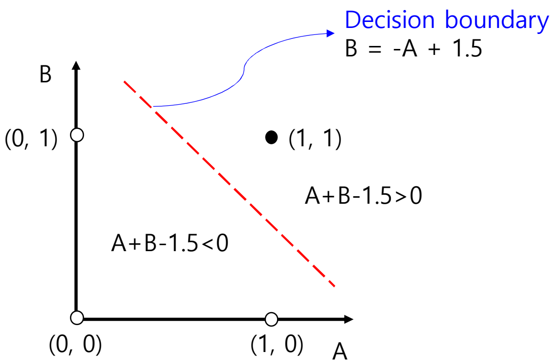
Logic Gates – AND

- Decision boundary of AND

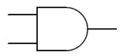


AND

A	B	Output
0	0	0
0	1	0
1	0	0
1	1	1

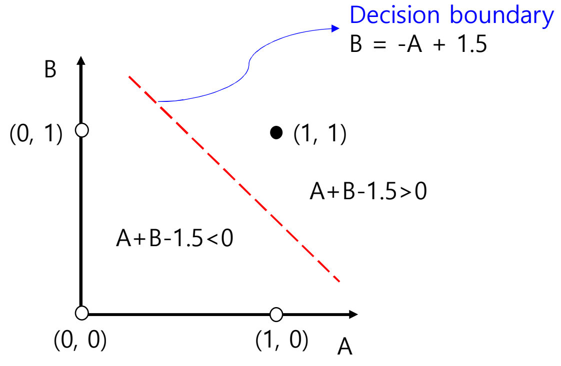


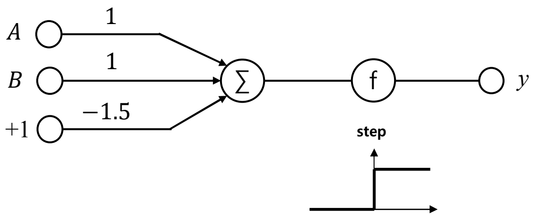
Logic Gates – AND



AND

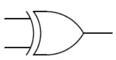
A	B	Output
0	0	0
0	1	0
1	0	0
1	1	1





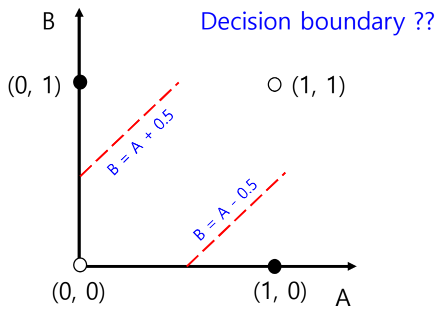
Logic Gates – XOR

- Decision boundary of XOR



XOR

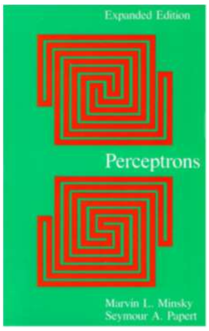
A	B	Output
0	0	0
0	1	1
1	0	1
1	1	0



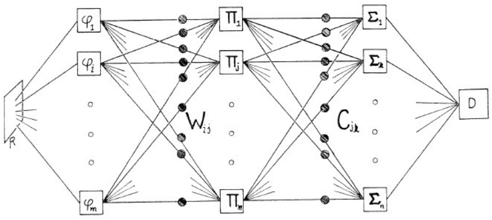
Rosenblatt's single-layer perceptron cannot classify input patterns that are not linearly separable.

MLP Can Do IT, But …

- Minsky and Papert, 1969



[206] **12.4** Learning Theory



It ought to be possible to devise a training algorithm to optimize the weights in this using, say, the magnitude of a reinforcement signal to communicate to the net the cost of an error. We have not investigated this.

[256] Epilogue

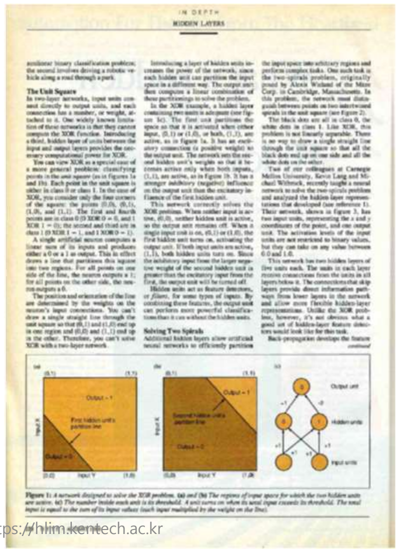
the above quotation as saying that, until recently, connectionism had been paralyzed by the following dilemma:

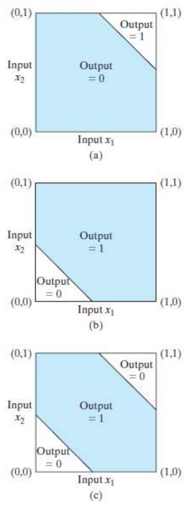
Perceptrons could learn anything that they could represent, but they were too limited in what they could represent.

Multilayered networks were less limited in what they could represent, but they had no reliable learning procedure.

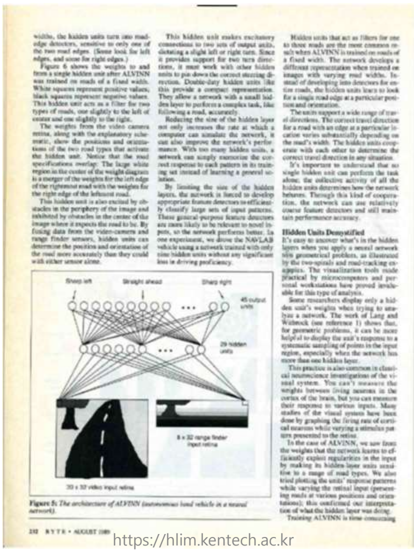
Logic Gates – XOR

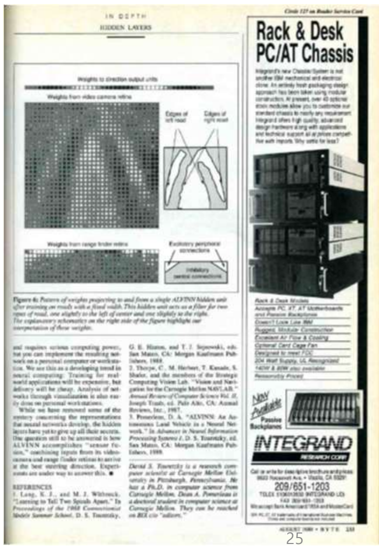
Touretzky and Pomerleau, 1989



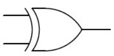


Logic Gates – XOR





Classification



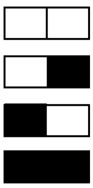
XOR

A	B	Output
0	0	0
0	1	1
1	0	1
1	1	0

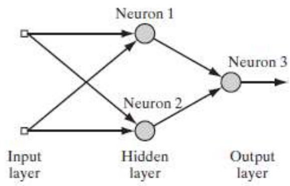
Pattern (2-bit) Class (0/1)

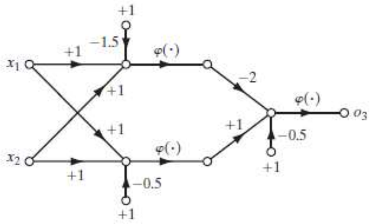
=

Pattern (2-bit) Class (0/1)

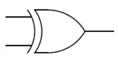


0
1
1
0





Classification

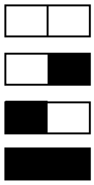


A	B	Output
0	0	0
0	1	1
1	0	1
1	1	0

Pattern (2-bit) Class (0/1)

=

Pattern (2-bit) | Class (0/1)



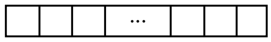
0

1

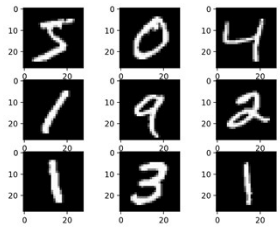
1

0

Pattern (k-bit)

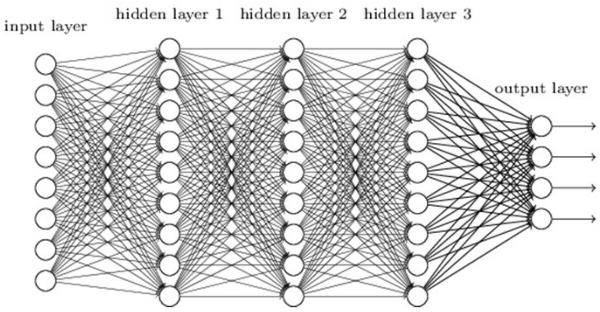


Pattern (28x28 grayscale image)

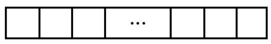


MNIST dataset

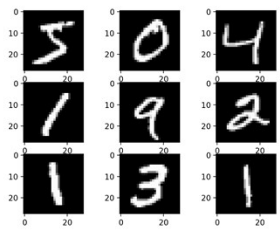
Classification



Pattern (k-bit)

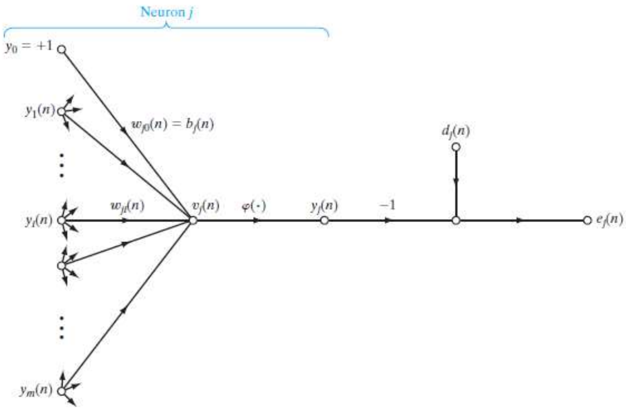


Pattern (28x28 grayscale image)



MNIST dataset

The Back-Propagation Algorithm



Total Error

$$\mathscr{E}(n) = \sum_{j \in C} \mathscr{E}_j(n)$$
$$= \frac{1}{2} \sum_{j \in C} e_j^2(n)$$

By Chain Rule

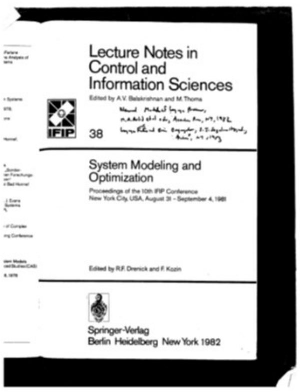
$$\frac{\partial \mathscr{E}(n)}{\partial w_{ji}(n)} = \frac{\partial \mathscr{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)}$$

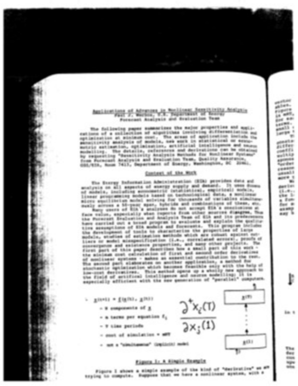
$$\frac{\partial \mathscr{E}(n)}{\partial w_{ji}(n)} = -e_j(n)\varphi_j'(v_j(n))y_i(n)$$

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathscr{E}(n)}{\partial w_{ji}(n)}$$

The Back-Propagation Algorithm

Paul John Werbos





Harvard Univ. 1974 / LNCIS 1982

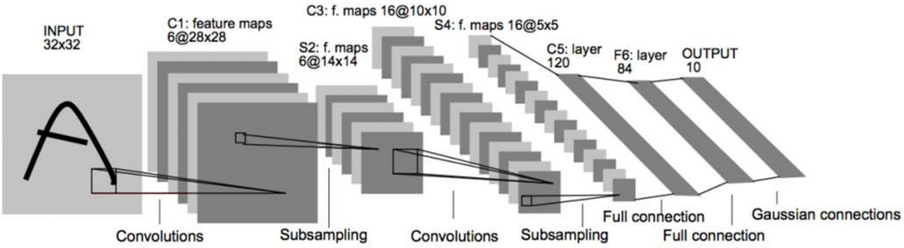
David Rumelhart, Geoffrey Hinton, Ronald Williams

Nature 1986

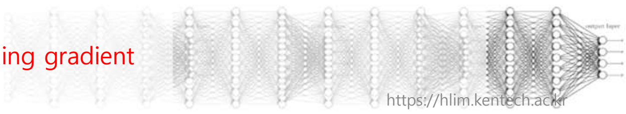
https://machinelearningknowledge.ai/timeline/finally-backpropagation-meets-neural-network/

LeNet-5 (1998)

- Yann LeCun, Leon Bottou, Yoshua Bengio, Patrick Haffner, "Gradient-Based Learning Applied to Document Recognition," Proc of The IEEE, 1998



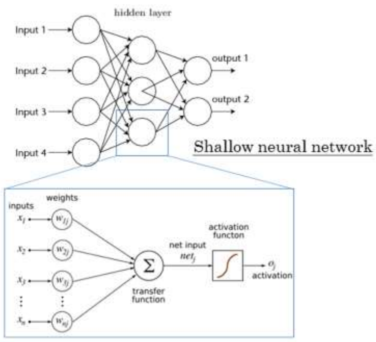
Vanishing gradient

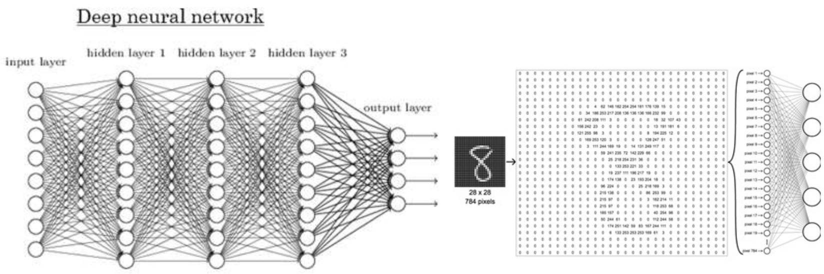


Research on other ML algorithms?

Deep Learning

Deep Neural Network (DNN)

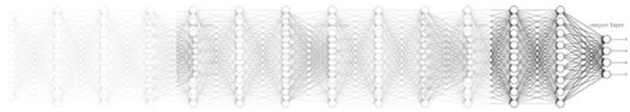




$net_j = x_1 w_{1j} + x_2 w_{2j} + \ldots$

$o_j = f(net_j)$

Vanishing gradient problem



Breakthrough

Big Data

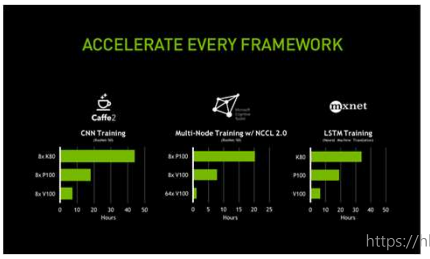
- Internet, SNS
- IoT

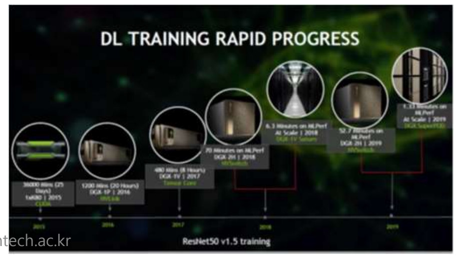
AI Algorithms

- DNN, CNN, RNN, …
- Open Source

Computing Power

- GPU, Cloud



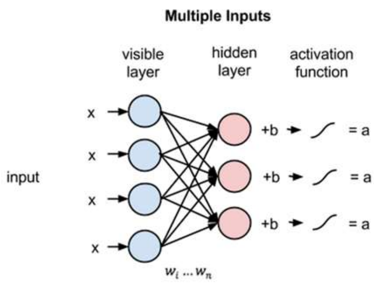


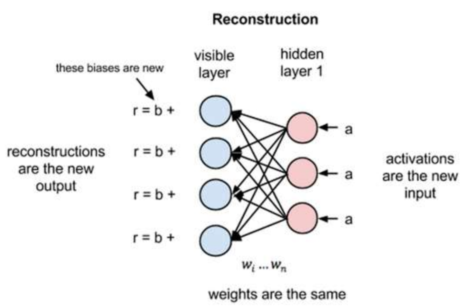
Why NOT Working Well in the Past?

- Geoffrey Hinton's Summary
  - Our labeled datasets were thousands of times too small.
  - Our computers were millions of times too slow.
  - We initialized the weights in a stupid way.
  - We used the wrong type of non-linearity.

Rebranded to "Deep Learning"

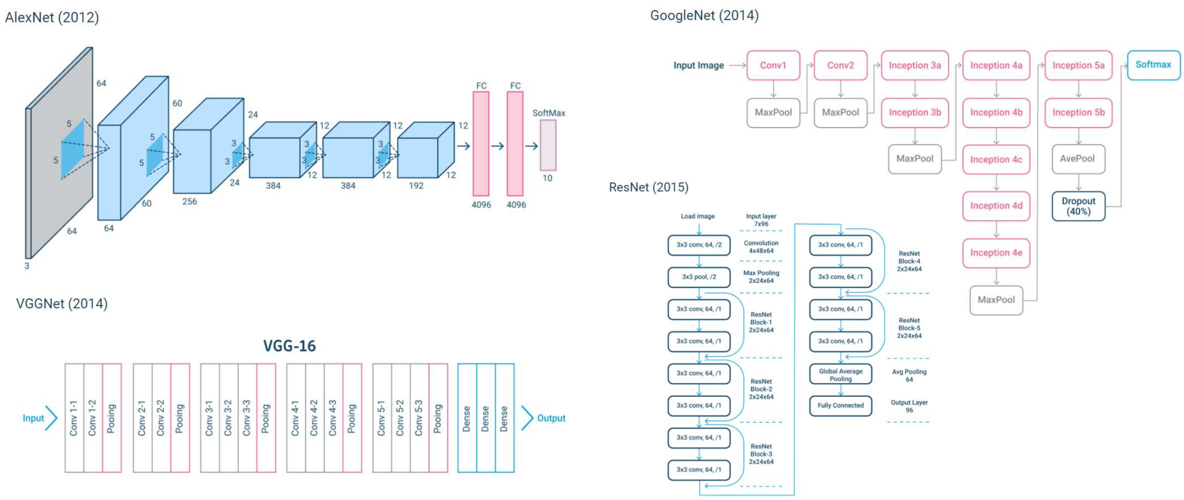
- "Rebrand" the frowned-upon field of neural nets with the moniker "deep learning"
- Hinton, Simon Osindero, and Yee-Whye The, "A fast learning algorithm for deep belief nets," Neural computation, 2006.





https://steemit.com/kr/@yoonheeseung/2-restricted-boltzmann-machines-rbms

Deep Learning for Computer Vision



Deep Learning for Natural Language Processing (NLP)

- OpenAI Generative Pre-trained Transformer 3 (GPT-3) with 175 billion machine learning parameters