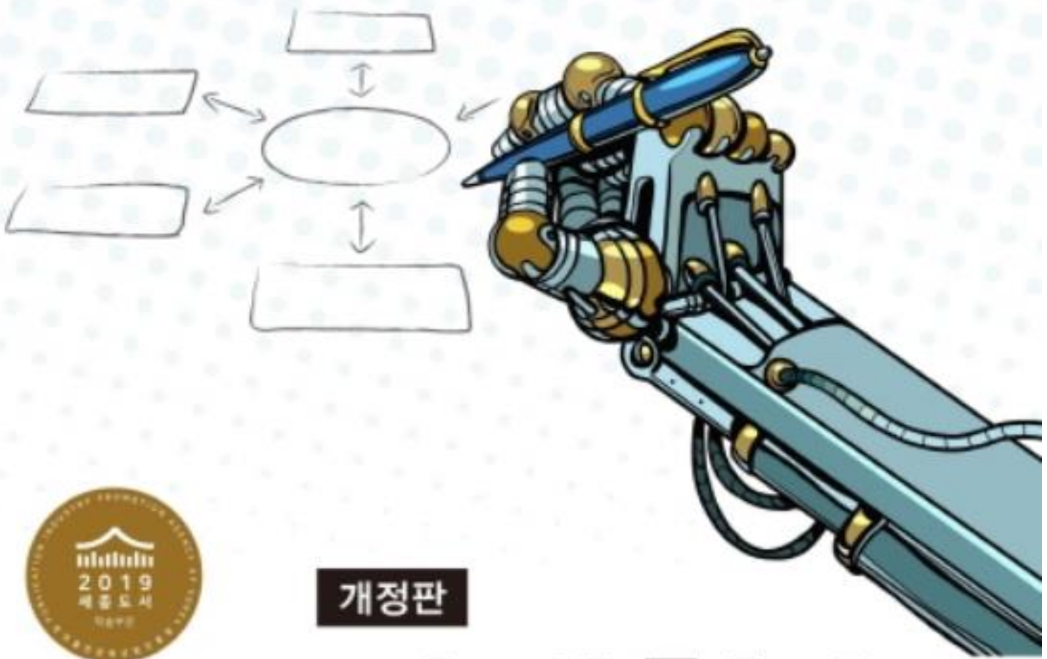
Machine Learning

Sang-Hoon Oh

Division of Information Communication Convergence Engineering

Mokwon University

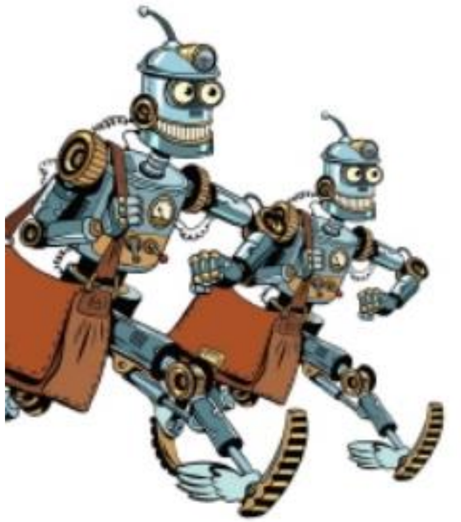


2019
세종도서

개정판

기계학습

수학적 이해에서 알고리즘까지



오상훈 지음

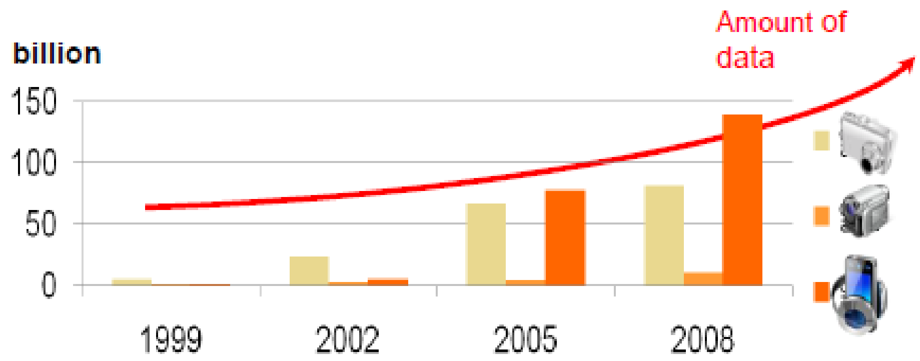
도서출판 홍릉

Contents

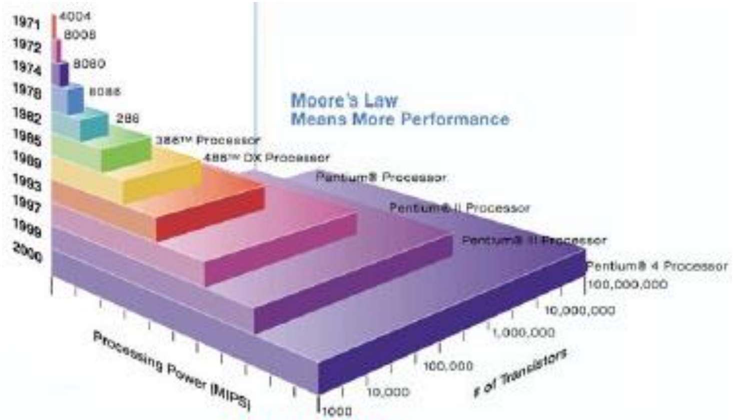
1.1. Machine Learning?

- Motivation for Machine Learning

- Advances in computer technology (store/access/process) can be used to analyze data (1) to make predictions, (2) for understanding and (3) control.
  - Moore's law : Describes a trend in the history of computer hardware. The number of transistors that can be inexpensively placed on an integrated circuit is increasing exponentially, doubling approximately every two years. The law is named after Intel co-founder Gordon E. Moore, who described the trend in his 195 paper. The term "Moore's law" was coined around 1970 by the Caltech professor, Carver Mead.
  - Data explosion : amount of data generated and stored grows at astounding rate (currently approximately 45GB/person x 7 billion = 315 Exabyte).



Data explosion

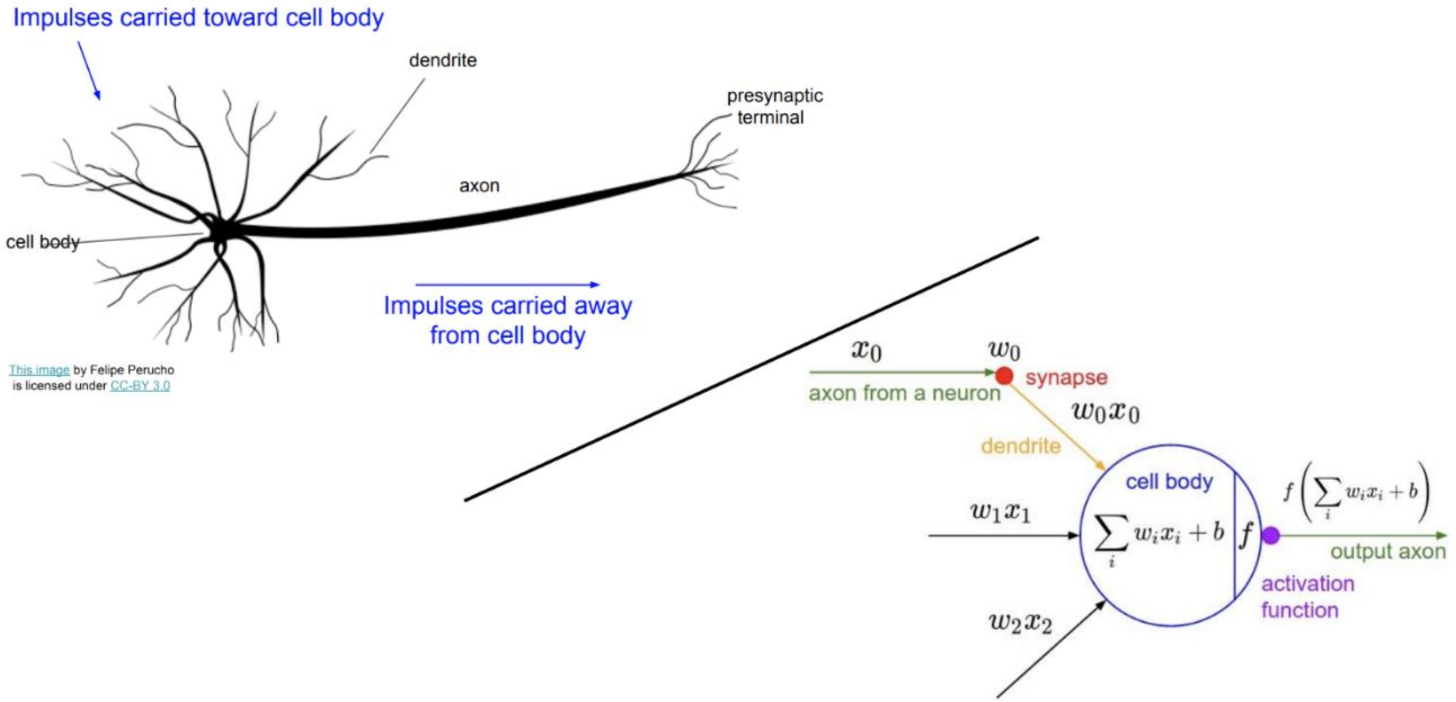


Moore's law

- May not understand the details of the process that explains the generation of the data (consumer behavior- buy beer with chips, ice cream in summer…), however, need a technique to make a good and useful approximation in detecting certain patterns and regularities. ⟹ Machine Learning!

1.2. Historical Background of Machine Learning

- Biological vs. Artificial Neuron
  - (1943) McCulloch and Pitts: Artificial neuron model
  - (1949) Hebbian learning
  - (1958) Perceptron (Rosenblatt)



Side-by-side illustrations of biological and artificial neurons, via Stanford's CS231n. This analogy can't be taken too literally—biological neurons can do things that artificial neurons can't, and vice versa—but it's useful to understand the biological inspiration. See Wikipedia's description of biological vs. artificial neurons for more detail.

1.2. Historical Background of Machine Learning

1. Latency Period(1940~1980)

Latency Period (1940~1980)	
1943	McClloch & Pitts Neuron Model
1949	Hebbian Learning (Hebb)
1958	Perceptron (Rosenblatt)
1960	Delta Rule (Widrow & Hoff)
1967	Outstar Learning (Grossberg)
1969	Perceptron Book (Minsky & Papert)
1972	Associative Memory Neural Nets (Kohonen)
1973	Pattern Classification and Scene Analysis (Duda & Hart)
1977	Associative Memory Nets (Anderson)
1980	Neocognitron (Fukushima)

2. Quickening Period(1980~1995)

Quickening Period (1980~1995)	
1981	Parallel Models of Associative Memory (Hinton & Anderson)
1982	Self-Organizing Maps (Kohonen)
1982	Hopfield Neural Networks (Hopfield)
1983	Boltzmann Machine (Hinton & Sejnowski)
1985	Adaptive Resonance Theory (Carpenter & Grossberg)
1986	Error Back-Propagation Algorithm (Rumelhart, Hinton, &Will iams)
1988	Bayesian Networks (Lauritzen, Spiegelhalter, & Pearl)
1992	Support Vector Machines (Boser, Guyon, & Vapnik)
1995	Statistical Learning Theory (Vapnik)

3. Growth Period(1995~2010)

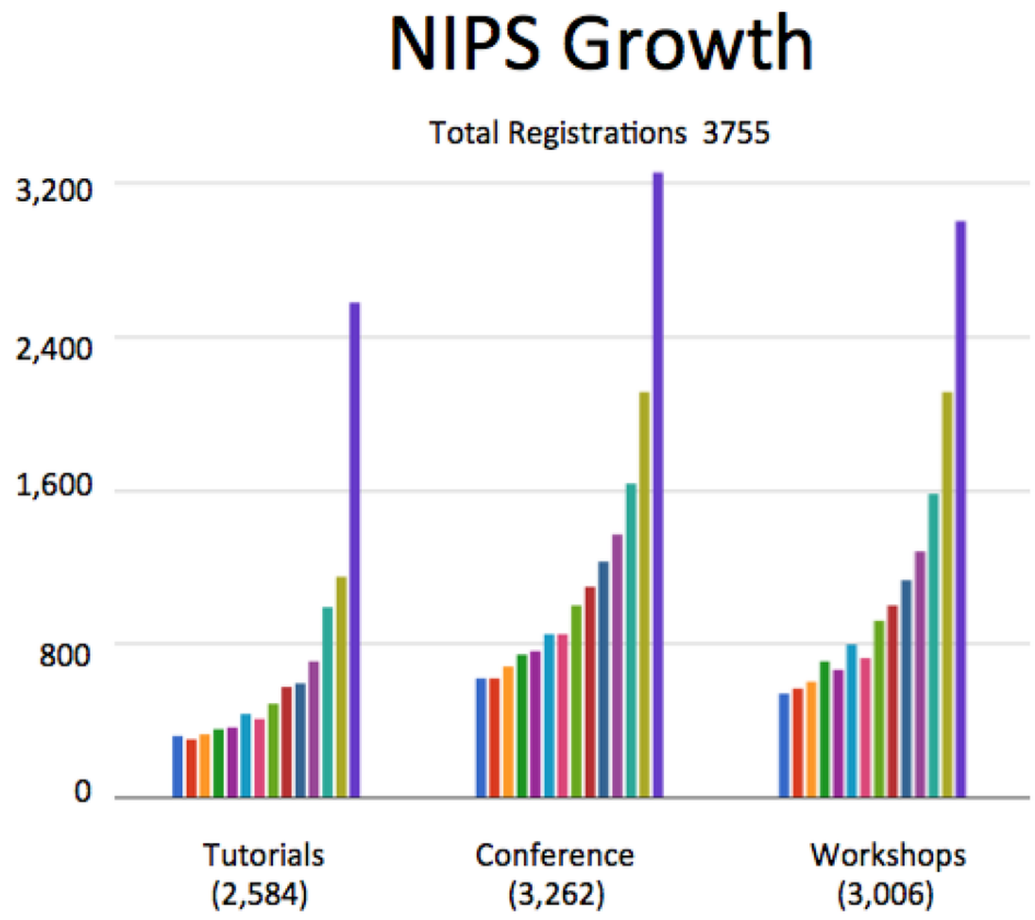
Growth Period (1995~2010)	
1997	Independent Component Analysis (Bell & Sejnowski)
1998	Natural Gradient (Amari)
1998	Reinforcement Learning (Sutton & Barto)
1999	Learning in Graphical Model (Jordan)
1999	Kernel Machines (Schoelkopf & Smola)
2003	Boosting Algorithms (Freund & Shapire)
2005	DARPA Grand Challenge
2006	Restricted Boltzmann Machine/Deep Learning (Hinton & Sal akutdinov)
2009	Probabilistic Graphical Models (Koller & Friedman)

4. Leaping Period(2010~Current)

Leaping Period (2010~Current)	
2010	Google Car
2011	IBM Watson AI
2011	Apple Siri Personal Assistant
2012	AlexNet
2013	Allen Institute for Artificial Intelligence
2013	Facebook AI Research
2014	Baidu Deep Learning Institute
2014	Amazon Echo & Alexa
2014	Generative Adversial Networks (Goodfellow et al.)
2015	DARPA Robotics Challenge
2016	Google DeepMind AlphaGo
2016	Apple AI Lab
2016	Google Home & Assistant

A Brief Review of NIPS 2015
February 2016
Article Authors: Zak Shafran

The twenty ninth Conference on Neural Information Processing Systems (NIPS), a single track machine learning and computational neuroscience conference, was held in Montreal, during a relatively balmy winter week, spanning December 7th to 12th. The conference saw a record number of attendees. As illustrated in the graph, during the last 14 years, there has been a steady increase in the number of attendees in all three categories -- tutorials, conference, and workshops; and biggest growth this year was in tutorials where the attendance doubled compared to the previous year!
....



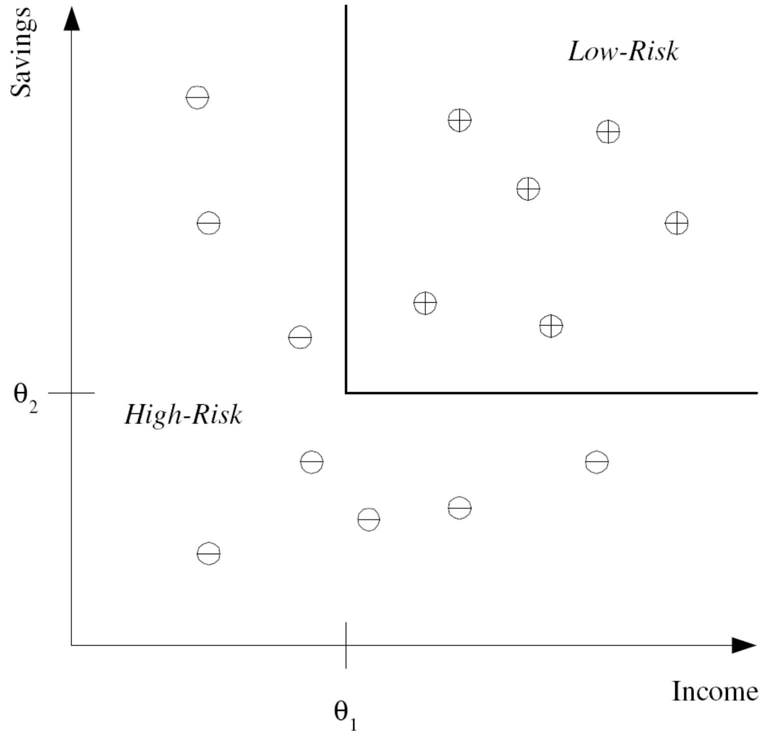
1.3. Types of Machine Learning

- Prediction
  - Supervised Learning: Given $\{(x_i, y_i)\}$ pairs, find a good understanding f: X→Y. Ex) Classification, Regression
  - Active Learning : Find f: X→Y, but $y_1, \ldots, y_N$ are initially hidden and there is a charge for each label you want revealed. The hope is that by intelligent adaptive querying, you can get away with significantly fewer labels than you would need in a regular supervised learning framework
  - Semi-supervised learning: Given both small number of labels and large number of unlabeled data, find f: X→Y.
- Understanding
  - Unsupervised Learning: Given $\{x_i\}$ find something interesting. Ex) Clustering, novelty-detection
- Policy(Control)
  - Reinforcement Learning: find f: P→A to max R

  (P: history of perception, A: action. R: long term reward)
- Generative Model
  - Generative Adversarial Network: The field of unsupervised learning which aims to study algorithms that learn the underlying structure of the given data, without specifying a target value.

Machine learning: **Classification**

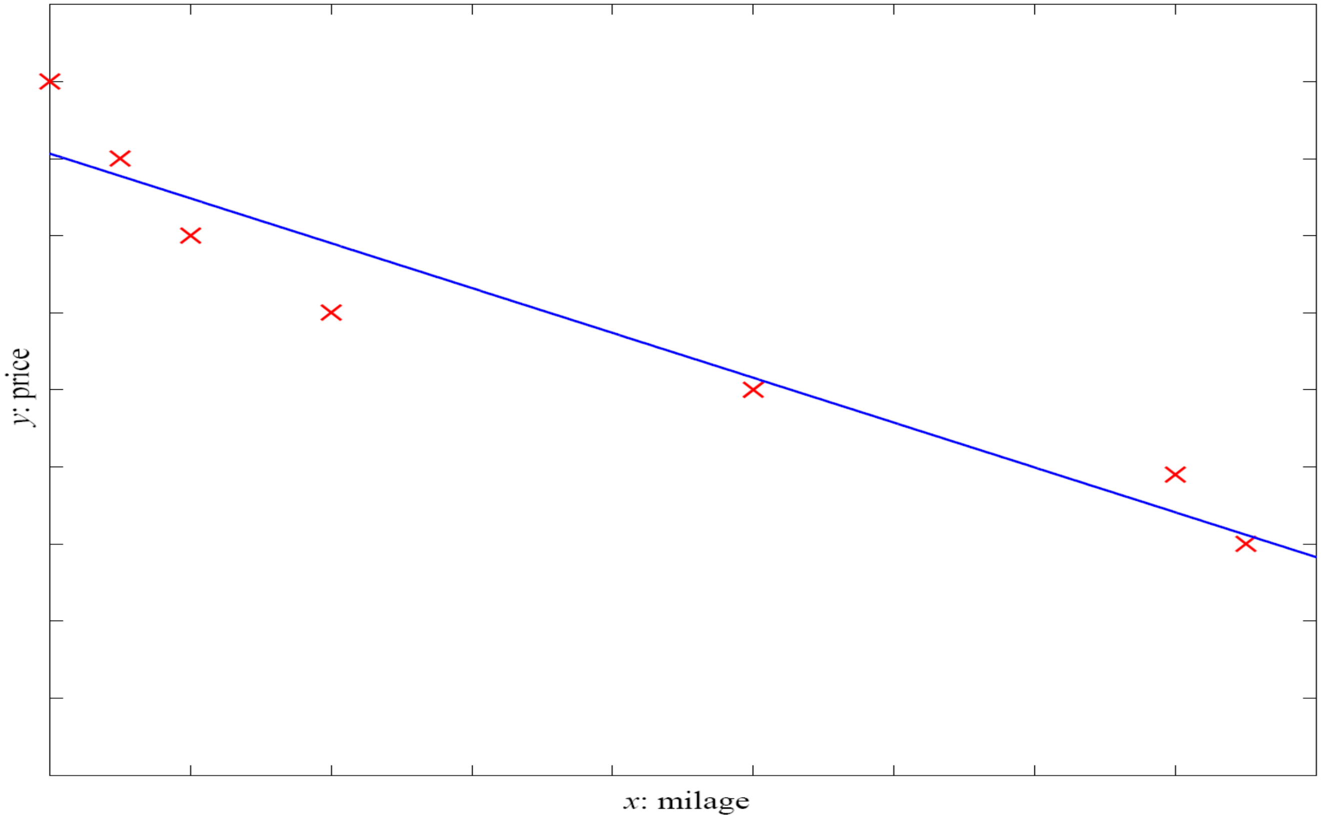
- Credit scoring
  - Differentiate between low-risk and high-risk customers from their incomes and savings
  - Learned classification rule:
    - IF income > θ_1 AND savings > θ_2 THEN low-risk ELSE high-risk



- Other application examples
  - Face recognition
  - Optical character recognition
  - Speech recognition
    - Input is temporal
    - Sensor fusion: integration of inputs from different modalities (e.g. acoustic and visual)
  - Outlier detection (fraud detection)

Machine learning: **Regression**

- Estimate the price of a used car
  - x = car attributes, y = price
  - y = g(x|θ)
    - g(·) is the model
    - θ is the parameter
  - Examples of g and θ
    - $g(x|w_0,w_1) = w_1 x + w_0$ (line
    - $g(x|w_0,w_1,w_2) = w_2 x^2 + w_1$ (quadratic model)
  - Output values continuous vs. classification (discrete)



- Other application examples
  - Autonomous car navigation
    - Learn the steer angle given input (video image, GPS, ...)
  - Typically have same application areas as classification

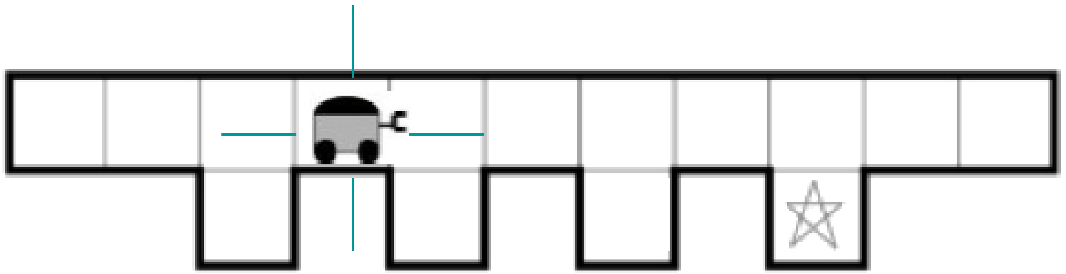
Machine learning: **Unsupervised Learning**

- ☐ Learning without correct output values (i.e., without supervisor)
  - vs. Classification & regression = supervised learning
  - Find regularities in the input
  - Statistics: density estimation
  - e.g. clustering (group similar instances)
- ☐ Example application
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs

Machine learning: **Reinforcement Learning**

- Learning a policy (a sequence of correct actions to reach the goal)
- No supervisor for telling the correct action: Learn from delayed reward

- Game playing
- Robot in the maze



- Multiple agents, partial observability, …

Applications

○**Medicine/Biology**

●(Predict) Patient hospitalized due to heart attack will have second attack based on demographic, diet and clinical measurements.

●(Predict) Glucose in the blood of diabetic person based on infrared absorption spectrum of person's blood.

●(Predict) Risk factor prostate cancer based on clinical and demographic variables).

●Bioinformatics.

○**Economy/finance**

●(Predict) Price of stock in 6 months based on company performance, economic data.

●(Predict) Credit card fraud based on gender, demographic, credit history

○**IT**

●(Predict) Numbers in a handwritten ZIP code based on digitized image of ZIP code

●(Predict) Spam mail based on text in mail

●(Predict) Text based on speech (Speech Recognition)

●(Understanding) Call patterns are analyzed for network optimization.

○**Astronomy/Physics**

●(Understanding) Discovering astronomical features

○**Intelligence**

●(Policy) Flying helicopters upside down based on altitude, wind factor …

○http://mmp.kaist.ac.kr/~ee531/invertedFlight.wmv

●(Policy) Playing game based on move by opponent.

●(Policy) Robot navigation based on sensor input.

Applications: Summary