



삼성 오픈소스 컨퍼런스
SAMSUNG OPEN SOURCE CONFERENCE

OPEN YOUR UNIVERSE WITH SOSCON

Searching for Magic Formula

MHR Inc.

안명호, mhr.james@gmail.com

2015-10-28



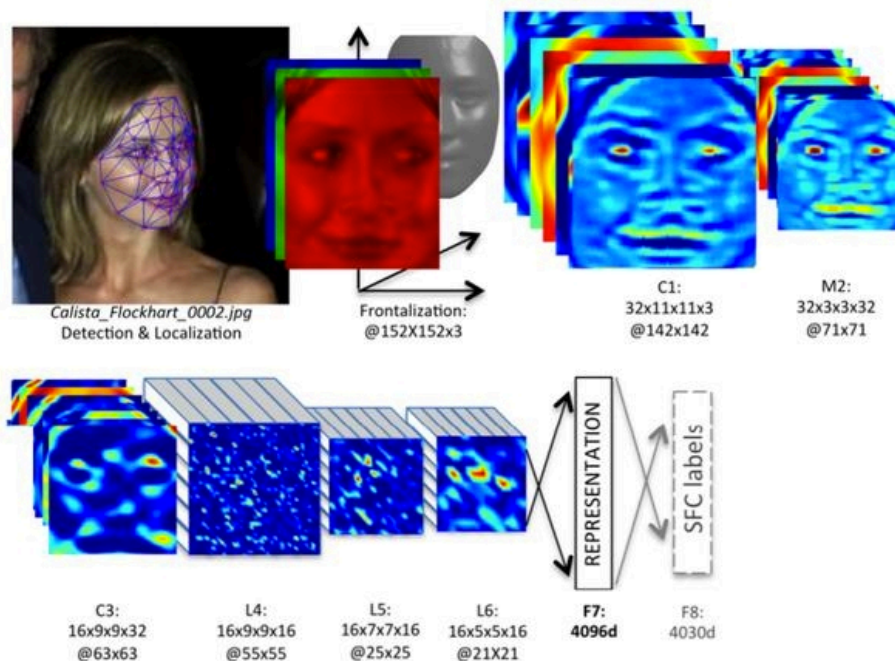
거대한 모험의 시작

블로터, 2014.07.31 이성규

페이스북은 딥러닝을 적용해 딥페이스라는 얼굴 인식 알고리즘을 올해 3월 개발했다.

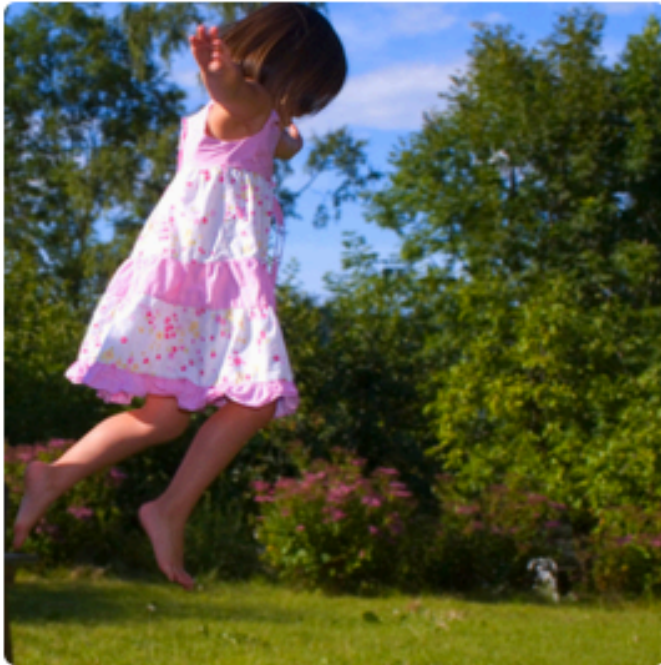
페이스북은 딥러닝이 적용된 딥페이스 알고리즘으로 전세계 이용자의 얼굴을 인식해 특정하고 있다.

인식 정확도는 97.25%로 인간 눈(97.53%)과 거의 차이가 없다. 페이스북은 이용자가 올린 이미지의 얼굴만 측면만 봐도, 어떤 이용자인지 판별해낼 수 있다.



source : <http://www.bloter.net/archives/201445>

Deep Learning을 이용한 Image Captioning 샘플



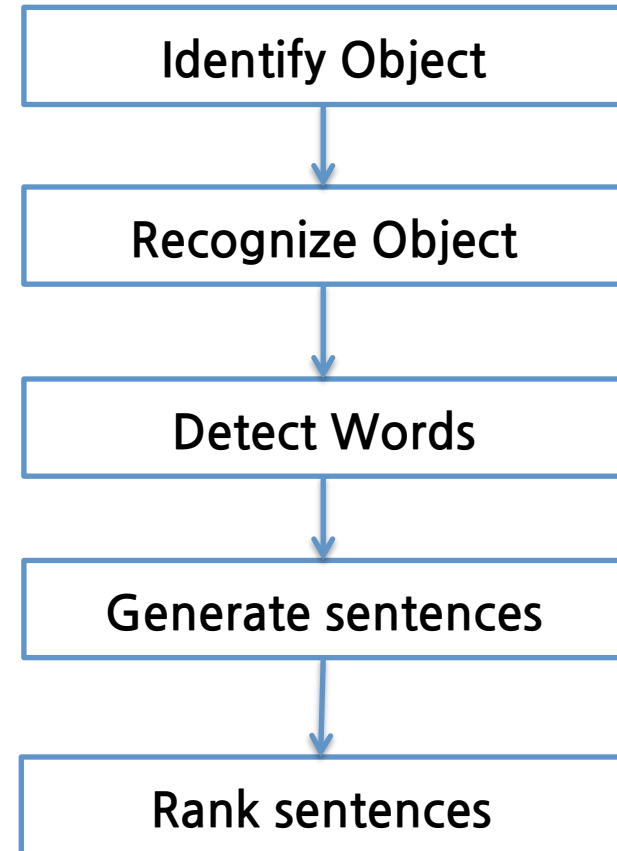
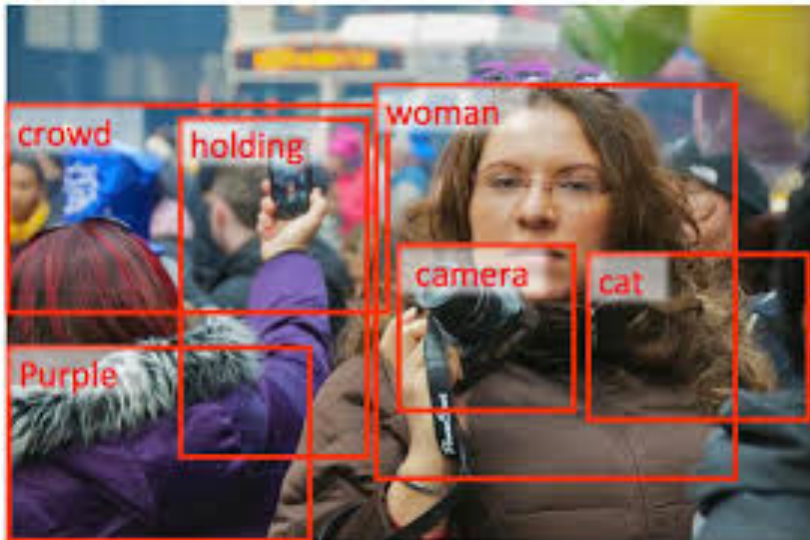
"girl in pink dress is
jumping in air."



"black and white dog
jumps over bar."

<http://cs.stanford.edu/people/karpathy/deepimagesent/>

결코 쉽지 않은 작업



The background is a deep blue gradient. A large, faint, light blue circle is centered on the page. Inside this circle, there is a dense pattern of concentric, slightly irregular lines that create a ripple effect. At the very center of these lines is a bright, multi-pointed starburst or nebula-like pattern in a lighter shade of blue. The overall effect is a sense of depth and focus towards the center.

Role Model

이데일리 2014.08.13, 이승현기자

사이먼스 회장은 미국 하버드대 수학과 교수에서 돌연 월가에서 헤지펀드(르네 상스테크놀로지) 회사 창립자로 변신, 총 재산이 13조원을 훌쩍 넘는 세계 88위의 부호로 성공했다.

그의 회사는 수학적 모델을 이용한 계량분석 기법인 '퀀트펀드'를 통해 지난 2000년대 연이어 대박행진을 이어갔다.

<http://www.edaily.co.kr/news/NewsRead.edy?SCD=JE41&newsid=02797846606186992&DCD=A00504&OutLnkChk=Y>



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

- Renaissance Technology 설립자
- MIT, UC 버클리 수학박사
- 전 하버드대 수학교수
- 수학자 및 헤지펀드 매니저
- 1976년 베블런상 수상
- 2006년 역대 최고 연봉
 - 17억달러(2조원)

실제로 그를 만나본 사람들은

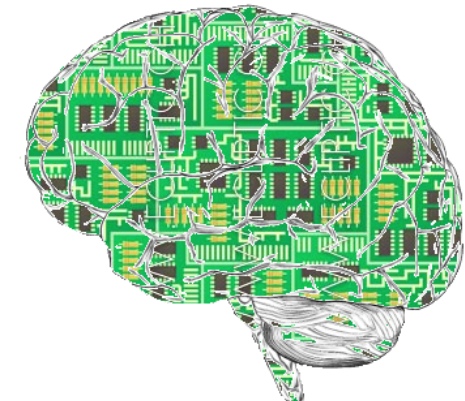
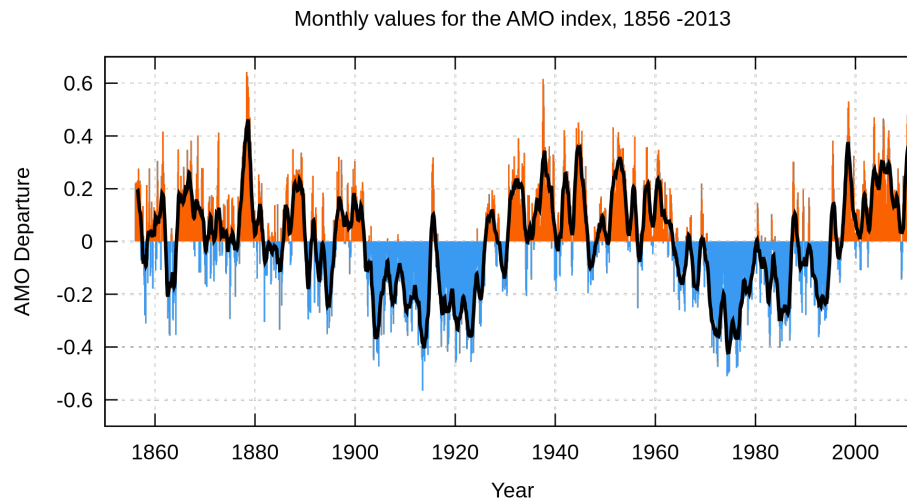
'사이먼스는 돈이 중요한 사람이라기보다는 단순히 수학에 미친 사람'

전통적인 금융회사와는 너무도 다른 금융회사



- 펀드관리 규모 \$22B
- 메달리온 펀드 연평균 36% 수익율
 - 워렌 버핏 연평균 수익율 21.6%
- 수학, 물리학, 천문학, 통계학자등 구성
- 전통적인 기업분석이나 주식분석 안함
- 알고리즘 트레이딩(Algorithm Trading)

결론은 패턴!!! 그렇다면 Deep Learning



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

Python library for Deep Learning

The screenshot shows a web browser window with the URL `deeplearning.net`. The page is titled "Theano 0.7 documentation" and has navigation links for "next", "modules", and "index". The main content area is titled "Welcome" and describes Theano as a Python library for defining, optimizing, and evaluating mathematical expressions. It lists several features: tight integration with NumPy, transparent use of a GPU, efficient symbolic differentiation, speed and stability optimizations, dynamic C code generation, and extensive unit-testing and self-verification. Below the features, it mentions that Theano has been powering large-scale computational intensive scientific investigations since 2007. A "News" section lists recent updates, including support for CuDNN v3, CNMeM, and Theano 0.7 release. At the bottom, there is a video player showing a talk at SciPy 2010.

Theano 0.7 documentation »

next | modules | index

Welcome

Theano is a Python library that allows you to define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays efficiently. Theano features:

- **tight integration with NumPy** – Use `numpy.ndarray` in Theano-compiled functions.
- **transparent use of a GPU** – Perform data-intensive calculations up to 140x faster than with CPU. (float32 only)
- **efficient symbolic differentiation** – Theano does your derivatives for function with one or many inputs.
- **speed and stability optimizations** – Get the right answer for `log(1+x)` even when `x` is really tiny.
- **dynamic C code generation** – Evaluate expressions faster.
- **extensive unit-testing and self-verification** – Detect and diagnose many types of mistake.

Theano has been powering large-scale computationally intensive scientific investigations since 2007. But it is also approachable enough to be used in the classroom (IFT6266 at the University of Montreal).

News

- We added support for [CuDNN v3](#).
- We added support for [CNMeM](#) to speed up the GPU memory allocation.
- Theano 0.7 was released 26th March 2015. Everybody is encouraged to update.
- We support [cuDNN](#) if it is installed by the user.
- Open Machine Learning Workshop 2014 [presentation](#).
- Colin Raffel [tutorial on Theano](#).
- Ian Goodfellow did a [12h class with exercises on Theano](#).
- New technical report on Theano: [Theano: new features and speed improvements](#).
- [HPCS 2011 Tutorial](#). We included a few fixes discovered while doing the Tutorial.

You can watch a quick (20 minute) introduction to Theano given as a talk at [SciPy 2010](#)

theano

Table Of Contents

- Welcome
- News
- Download
- Status
- Citing Theano
- Documentation
- Community
- Help!
 - [How to Seek Help](#)
 - [How to provide help](#)

Next topic

Release Notes

This Page

Show Source

Quick search

Enter search terms or a module, class or function name.



Born to be a fin-tech eng
Maybe...

불행하게도... ..



내가 필요한 것들...

하지만 국내 IT 환경에서는 매우 얻기 힘든...

- 과로하지 하지 말것
- 스트레스 받지 말것
- 적절한 휴식과 운동

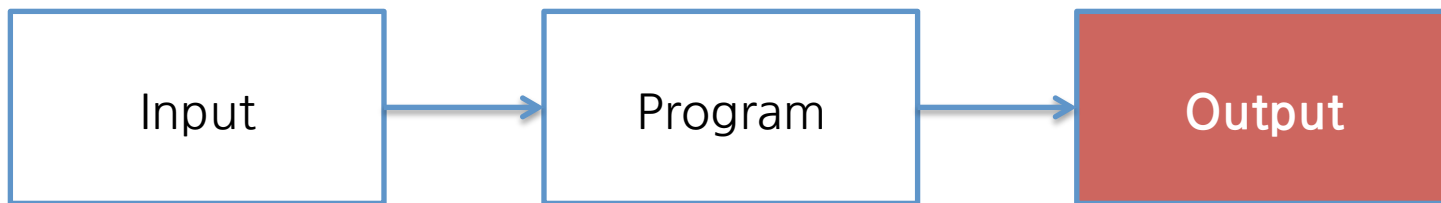




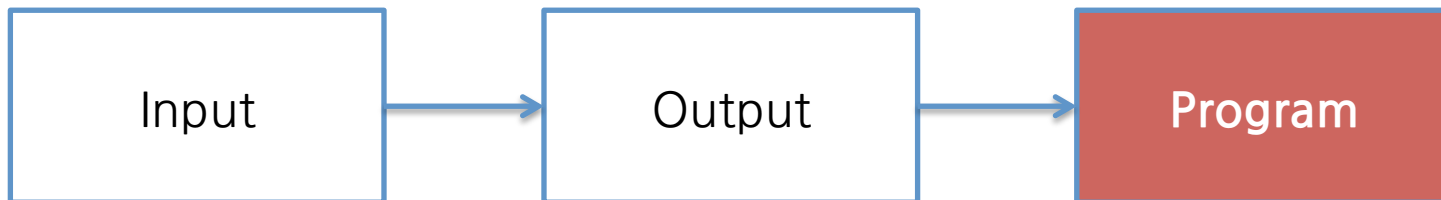
The background is a deep blue gradient. A large, faint, light-blue circle is centered on the page. Inside this circle, there is a starburst or nebula-like pattern of small, bright blue dots and lines radiating from a central point. The overall effect is a futuristic or technological aesthetic.

Machine Learning

Non-Machine Learning

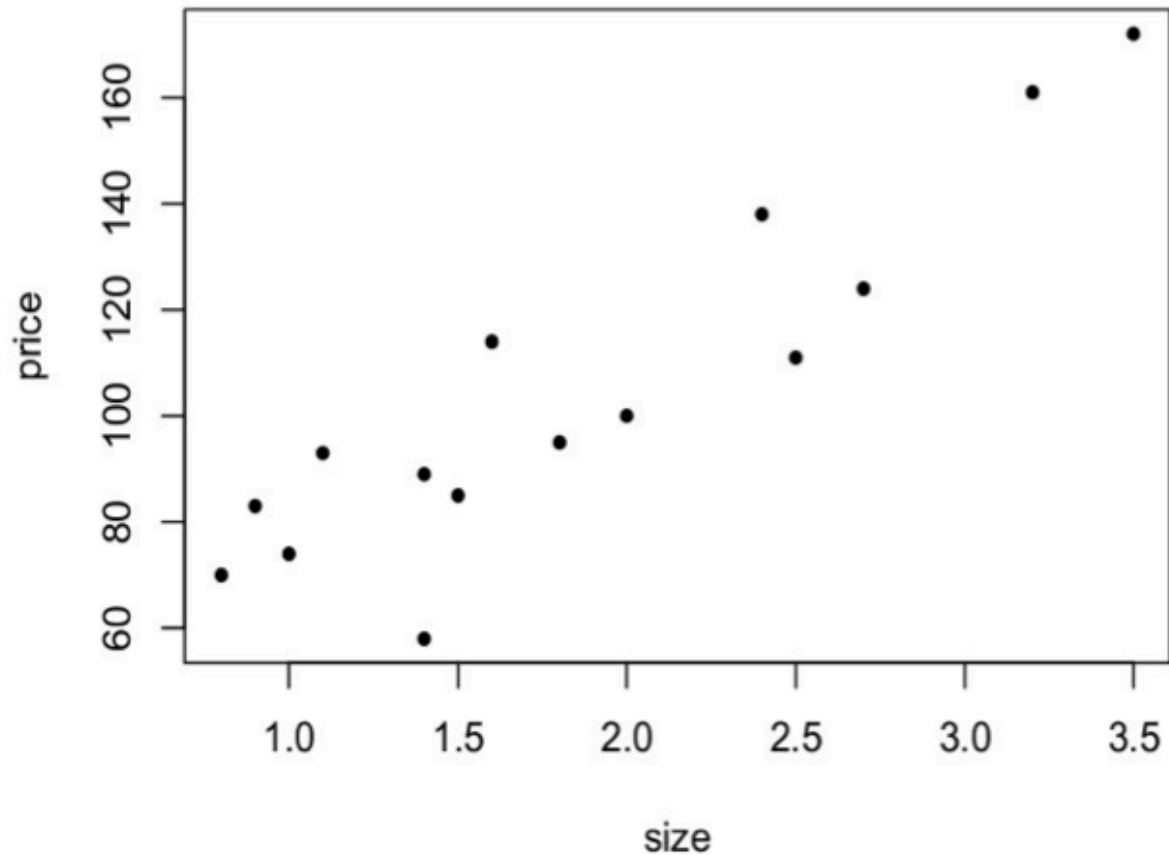


Machine Learning



아래처럼 주택가격 예측을 하고 싶다면 ...

Size (x)	Price (y)
0.80	70
0.90	83
1.00	74
1.10	93
1.40	89
1.40	58
1.50	85
1.60	114
1.80	95
2.00	100
2.40	138
2.50	111
2.70	124
3.20	172
3.50	172

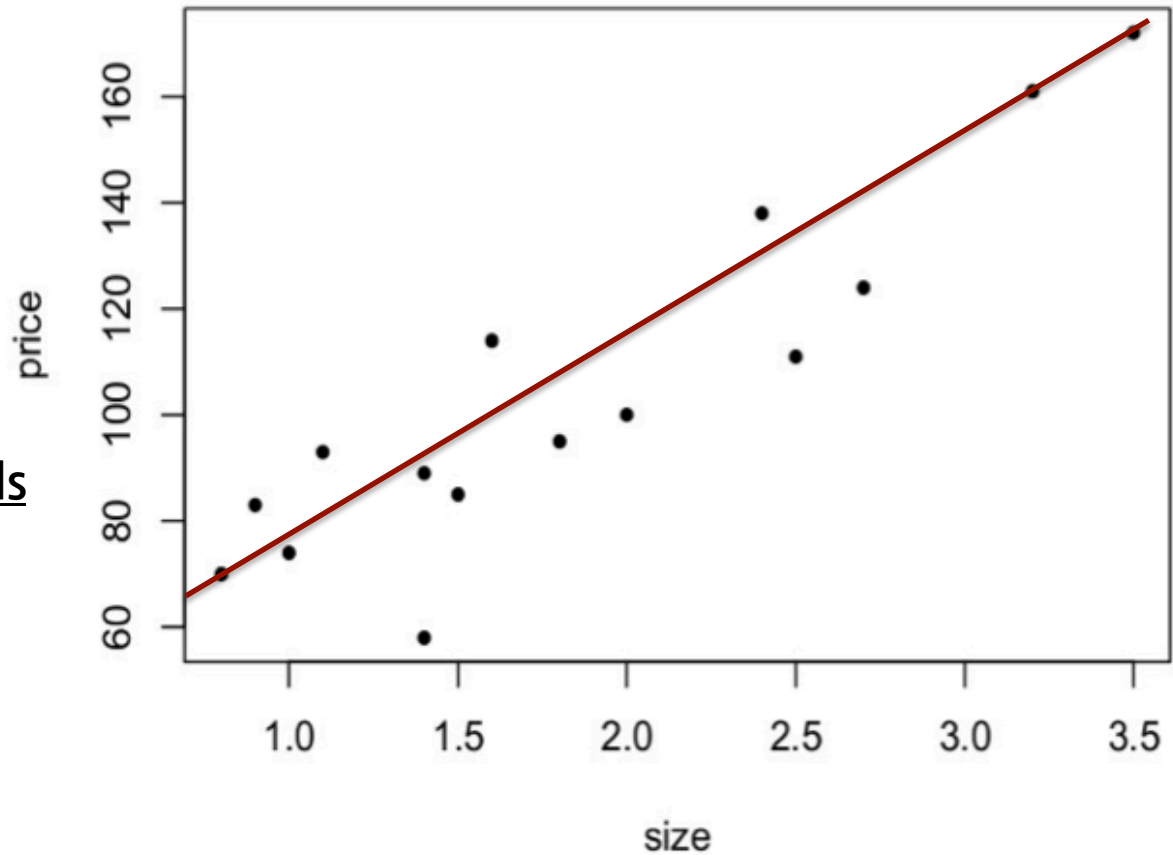


수많은 반복과 노력이 필요...

Hypothesis

$$Y = ax + b$$

Find 'a' and 'b' By hands

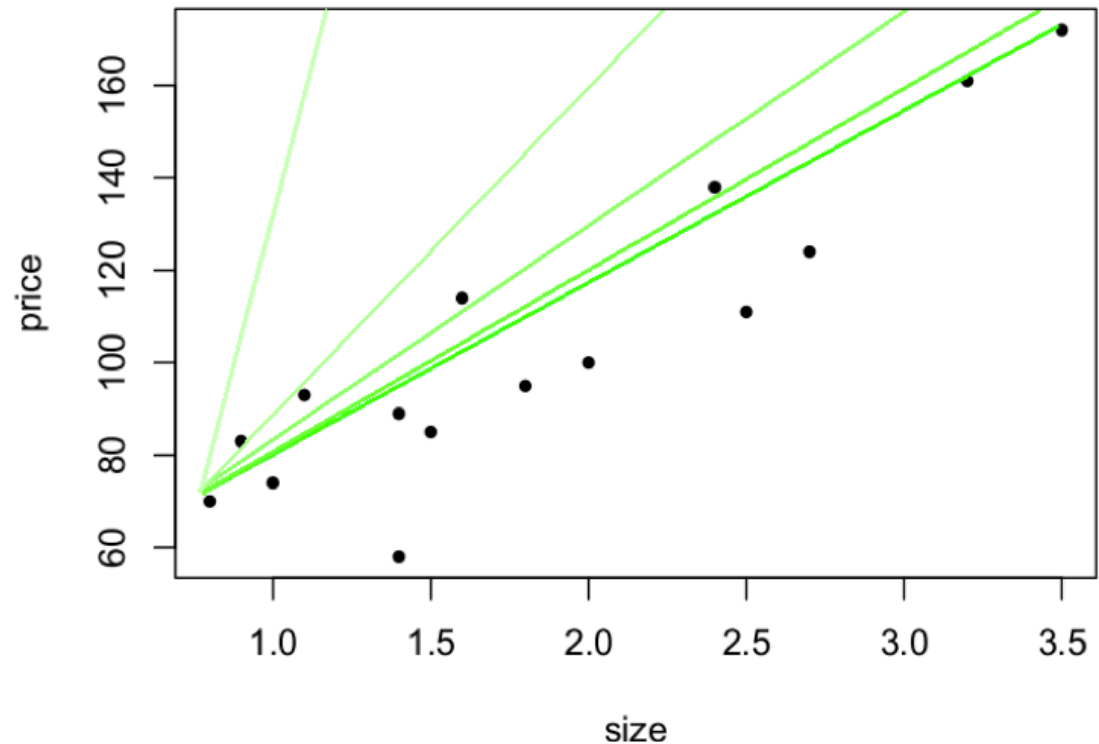


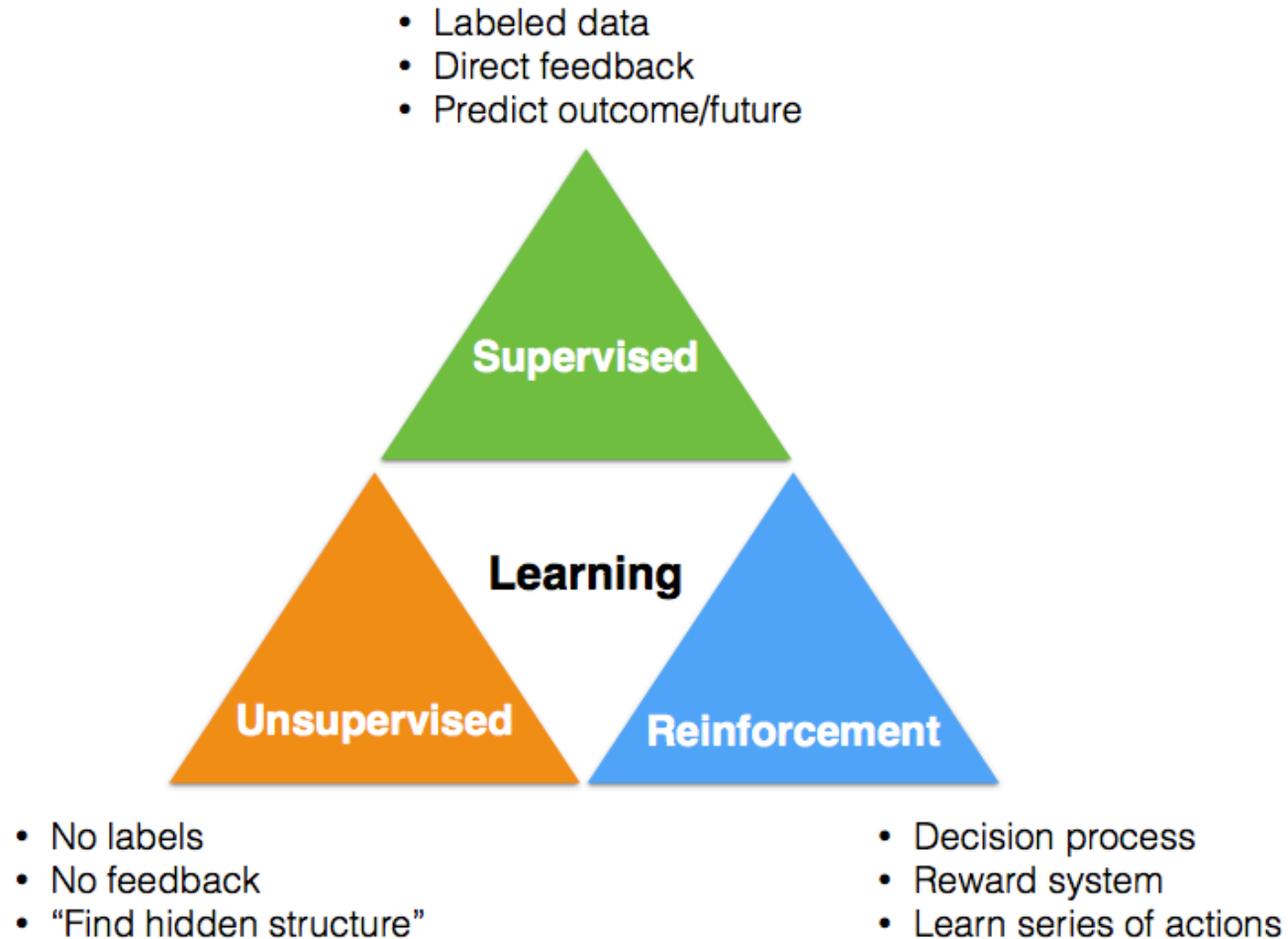
기계에 의한 수많은 반복

Hypothesis

$$Y = ax + b$$

Find 'a' and 'b' By Iteration





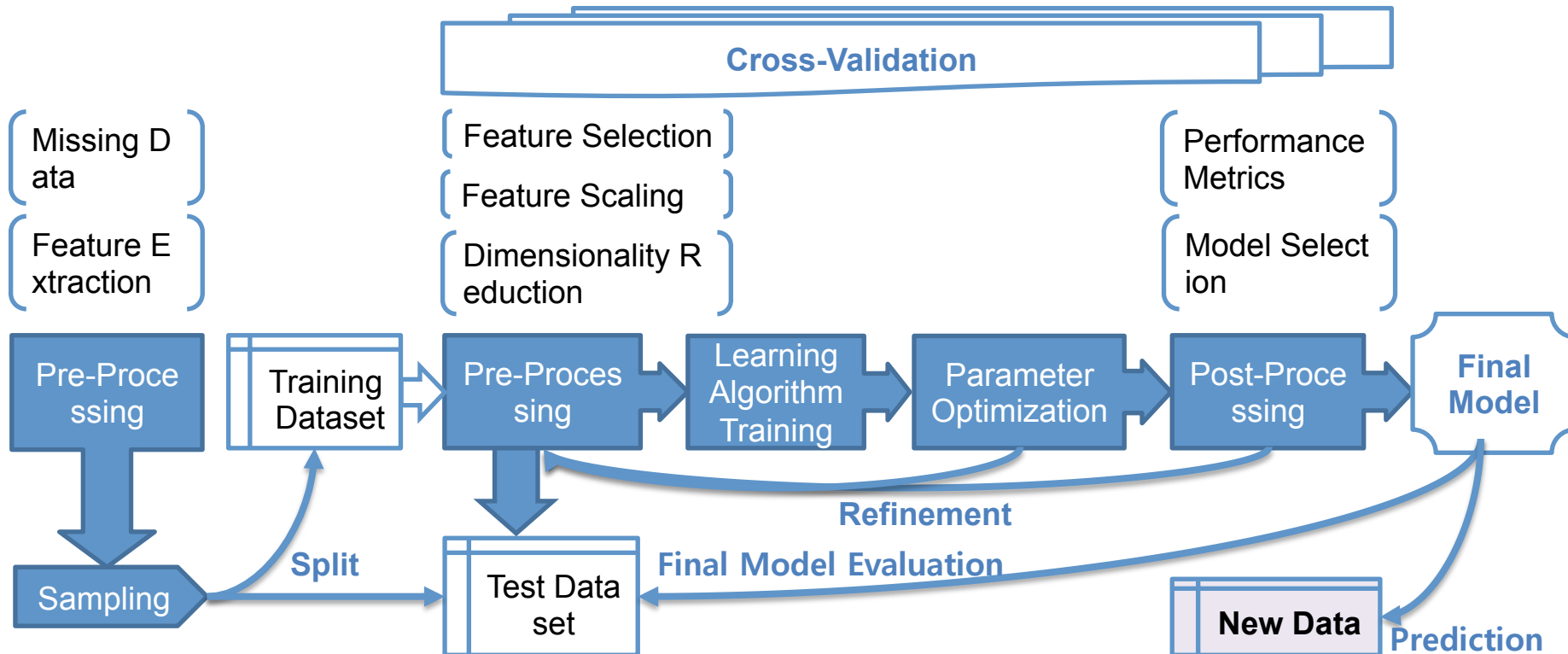


<http://dataaspirant.com/2014/09/19/supervised-and-unsupervised-learning/>



<http://dataaspirant.com/2014/09/19/supervised-and-unsupervised-learning/>

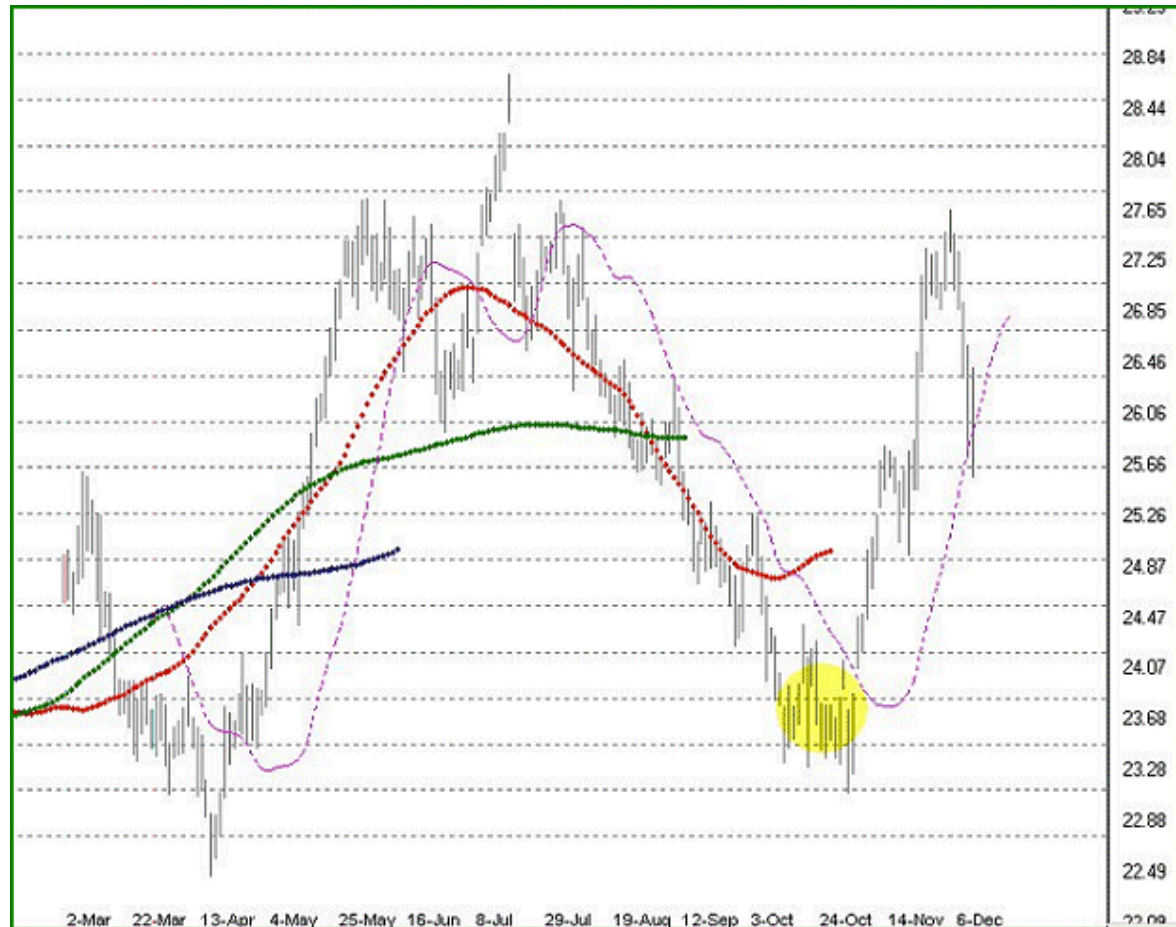
사실 알고리즘은 전체의 작은 부분이다.



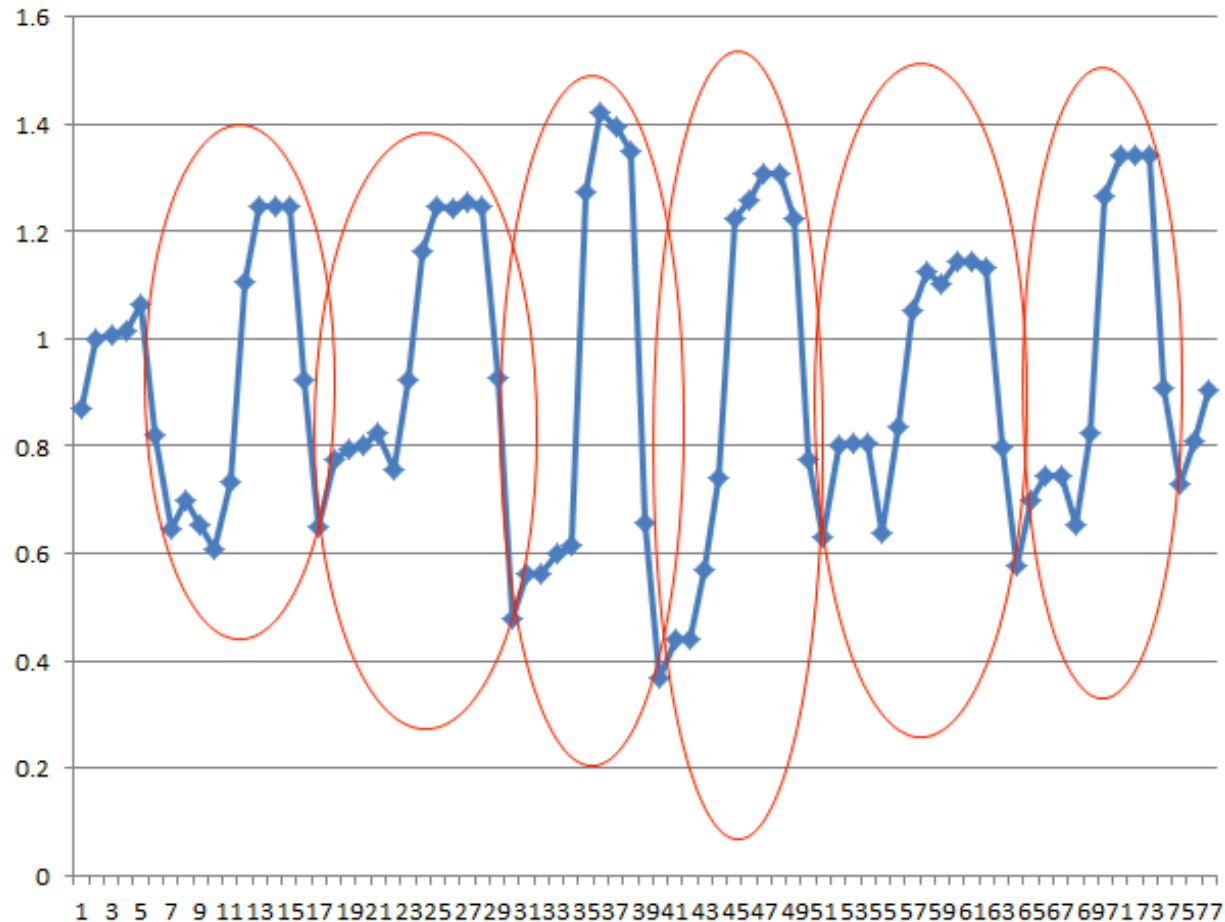


첫번째 시도

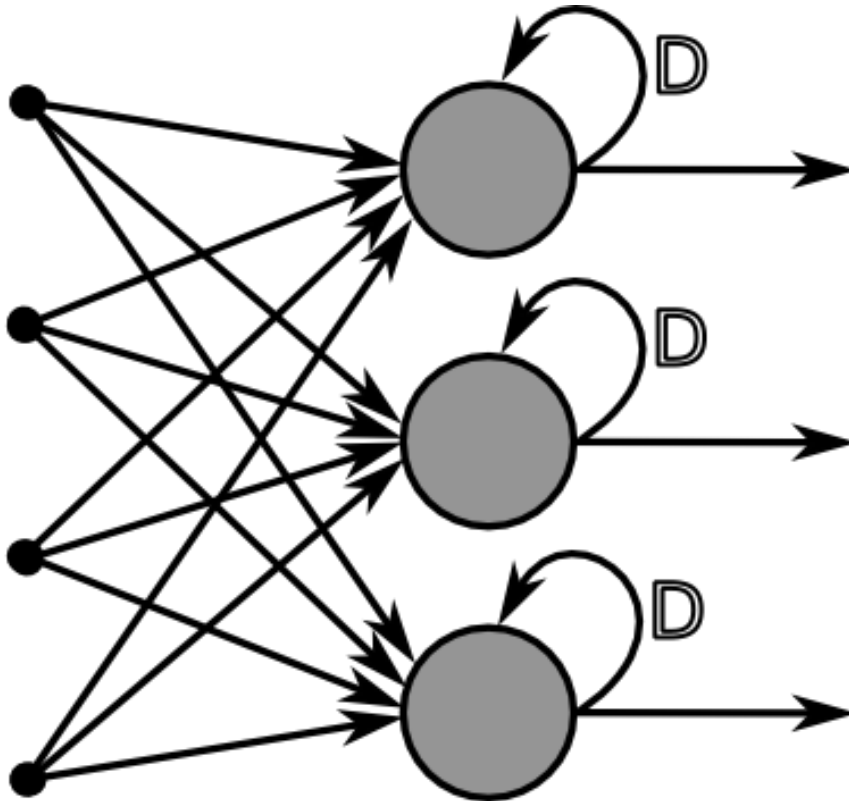
나도 알고리즘 트레이딩을 해보자.



Pattern 인식이 필요해, 정말로 ...



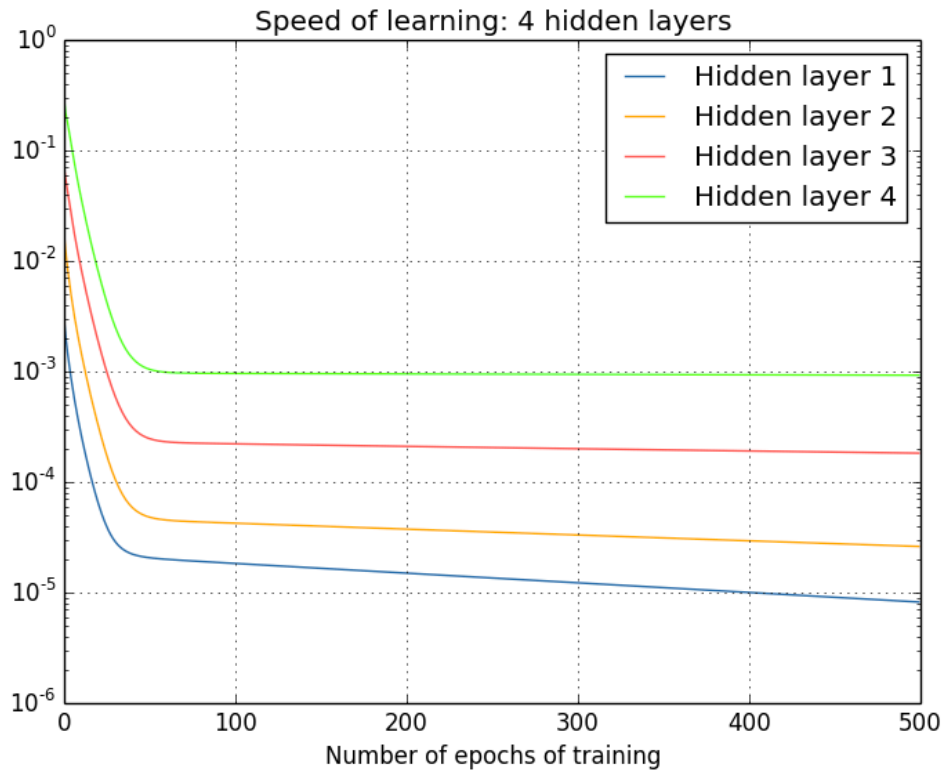
Time Series Data를 위한 Deep Learning 알고리즘



A recurrent neural network (RNN) is a class of artificial neural network where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior. Unlike feedforward neural networks, **RNNs can use their internal memory to process arbitrary sequences of inputs.** This makes them applicable to tasks such as unsegmented connected handwriting recognition, where they have achieved the best known results.

Source : wiki

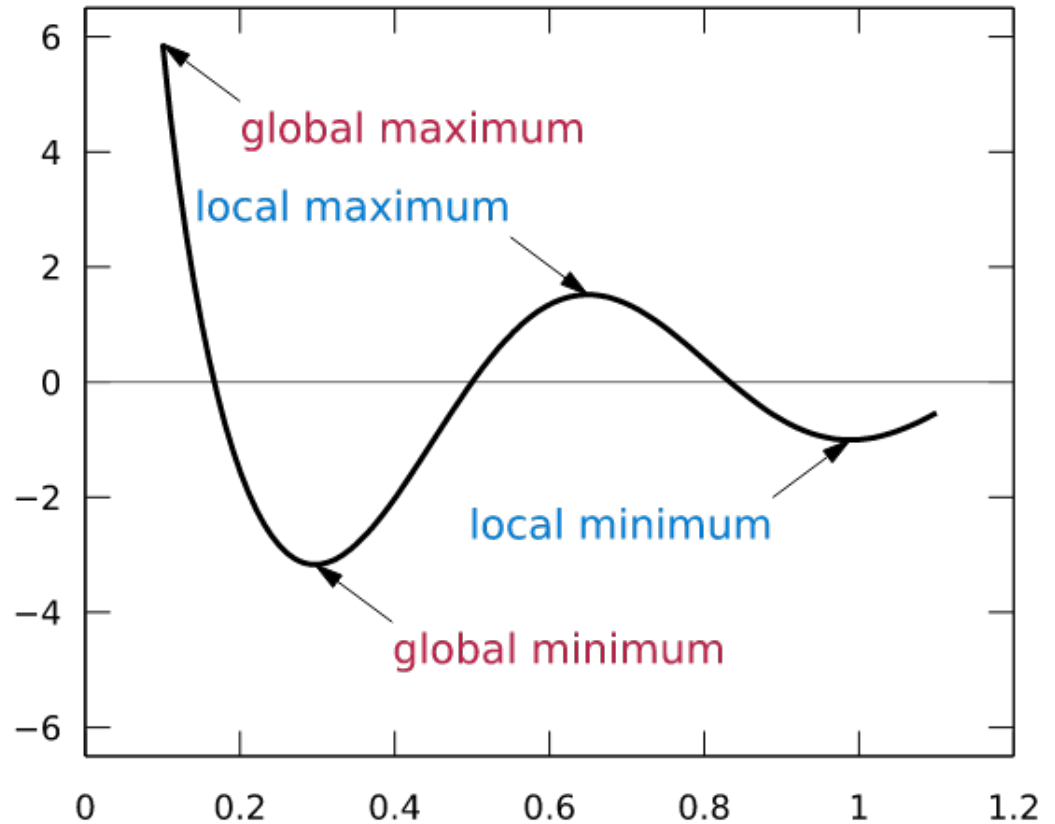
Vanishing Gradient Problem!!! Hard to train by Gradient-based training algorithm



1st Layer : Slowest

Last Layer : Fastest

간단히 이야기 하면 최적화 문제

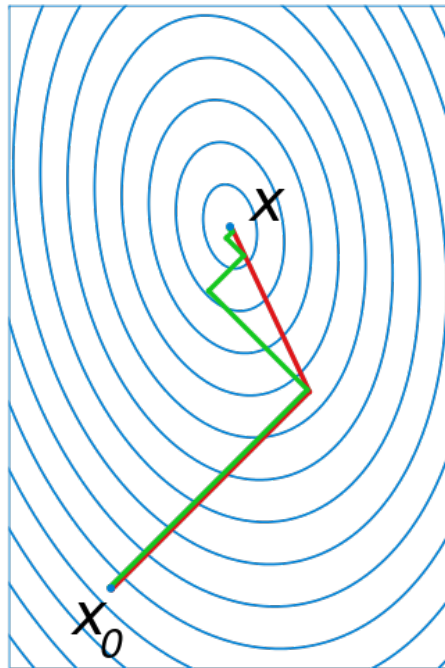


<http://www.toptal.com/machine-learning/an-introduction-to-deep-learning-from-perceptrons-to-deep-networks>

RNN문제를 해결하기 위해 나에게 도움을 주었던 논문

Learning Recurrent Neural Networks with Hessian-Free Optimization

By Martens & Sutskever



- Green : Gradient Descent
- Red : Conjugate Gradient

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \lambda_k \nabla f(\mathbf{x}_k), \quad k \geq 0$$

Gradient Descent

Learning rate 'a' is fixed

Conjugate Gradient

$$\alpha = - \frac{d_i^T (Ax_i + b)}{d_i^T A d_i}.$$

하지만 결론은 ...

알고리즘은 알고리즘일뿐...



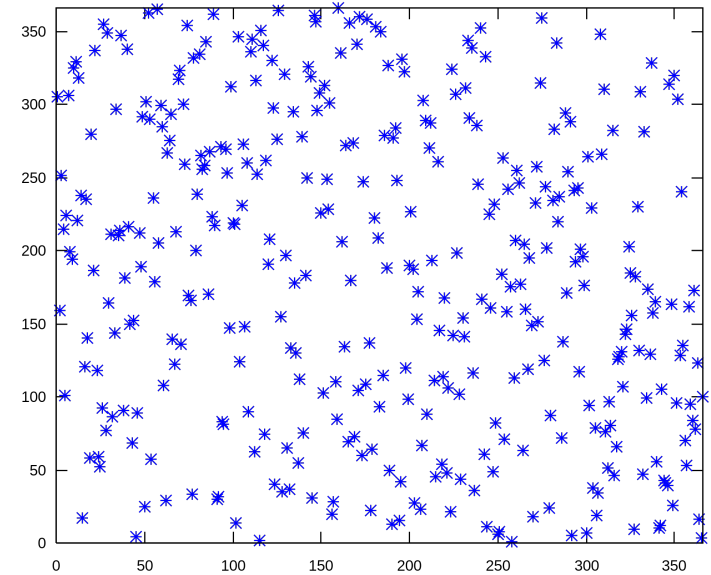
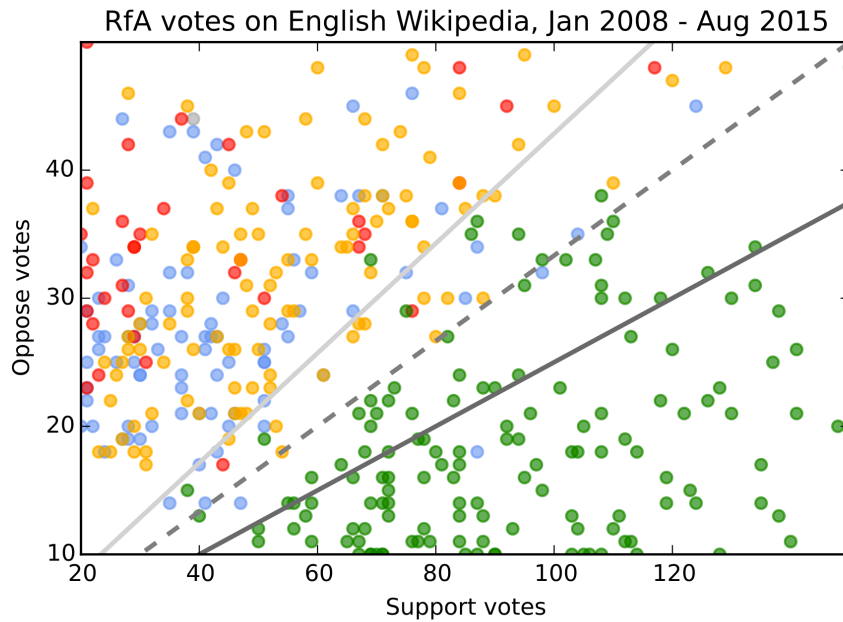


두번째 시도

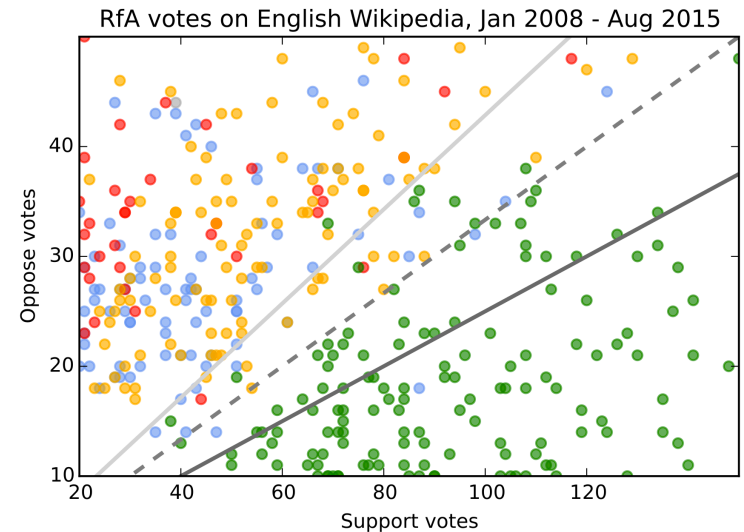
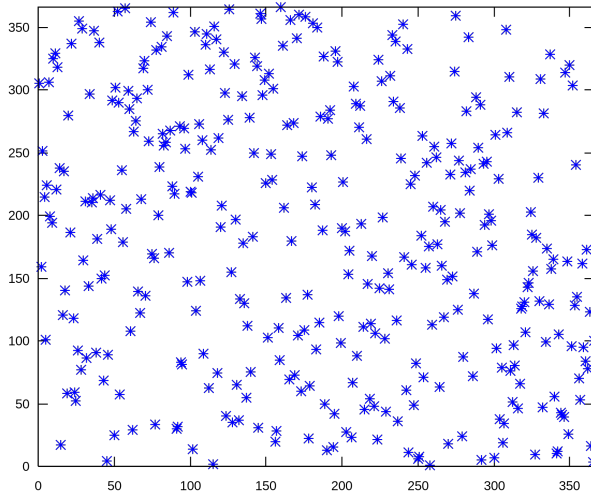
기본으로 돌아가자.
중요한 것은 데이터이다.!!!



패턴을 찾을 수 있는가?

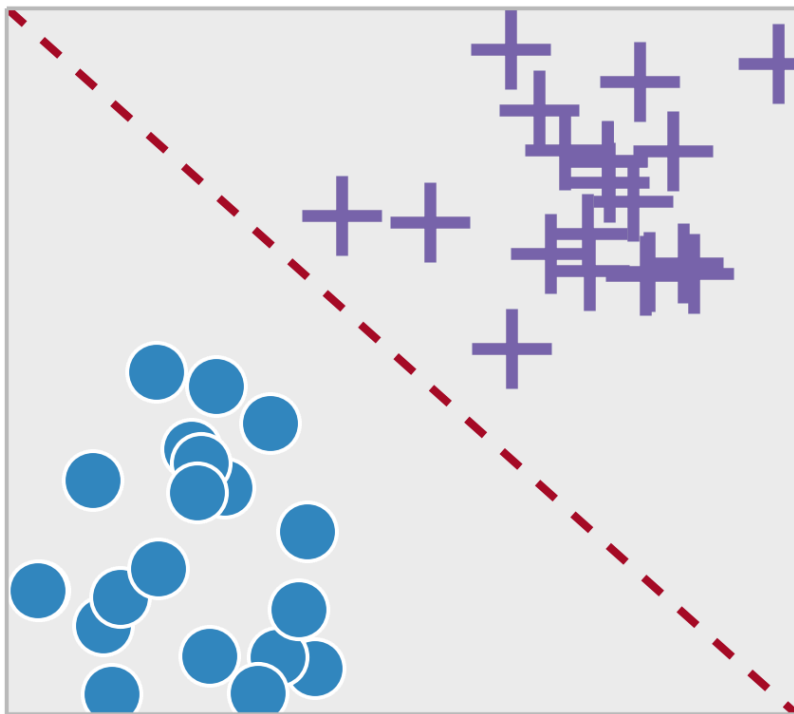


Data Preprocessing을 통해 예측가능한 형태로...
그래서 Domain Knowledge가 매우 중요!!!

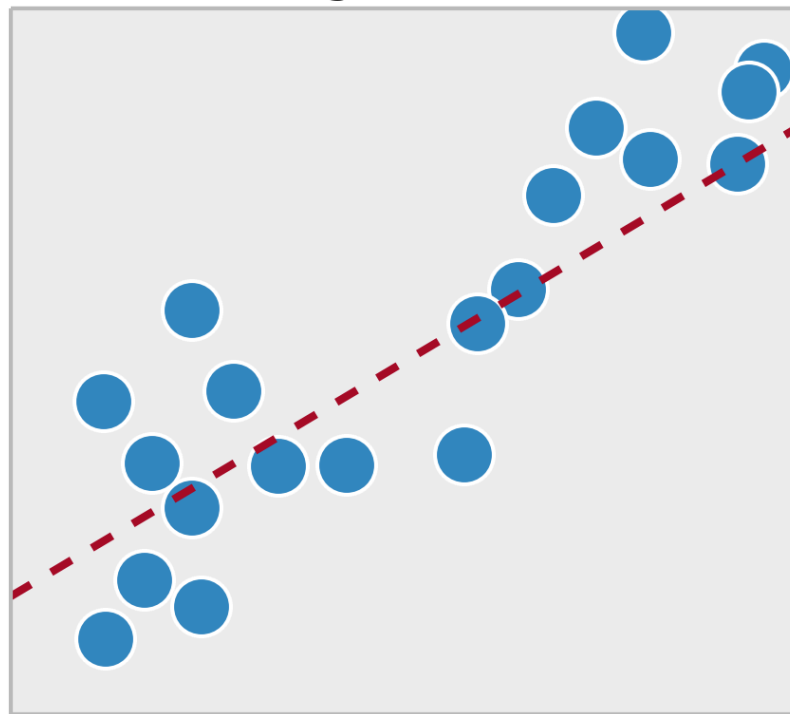


Regression Analysis 많으로도 충분

Classification



Regression



Deep Learning을 이용한 첫번째 시도보다 9배 향상



Machine Learning을 적용하기 위한 최적의 Domain?



23전 23승

선승구전(先勝求戰)

이기는 군대는 미리 이겨놓고 싸운다



삼성 오픈소스 컨퍼런스
SAMSUNG OPEN SOURCE CONFERENCE

OPEN YOUR UNIVERSE WITH SOSCON

THANK YOU!