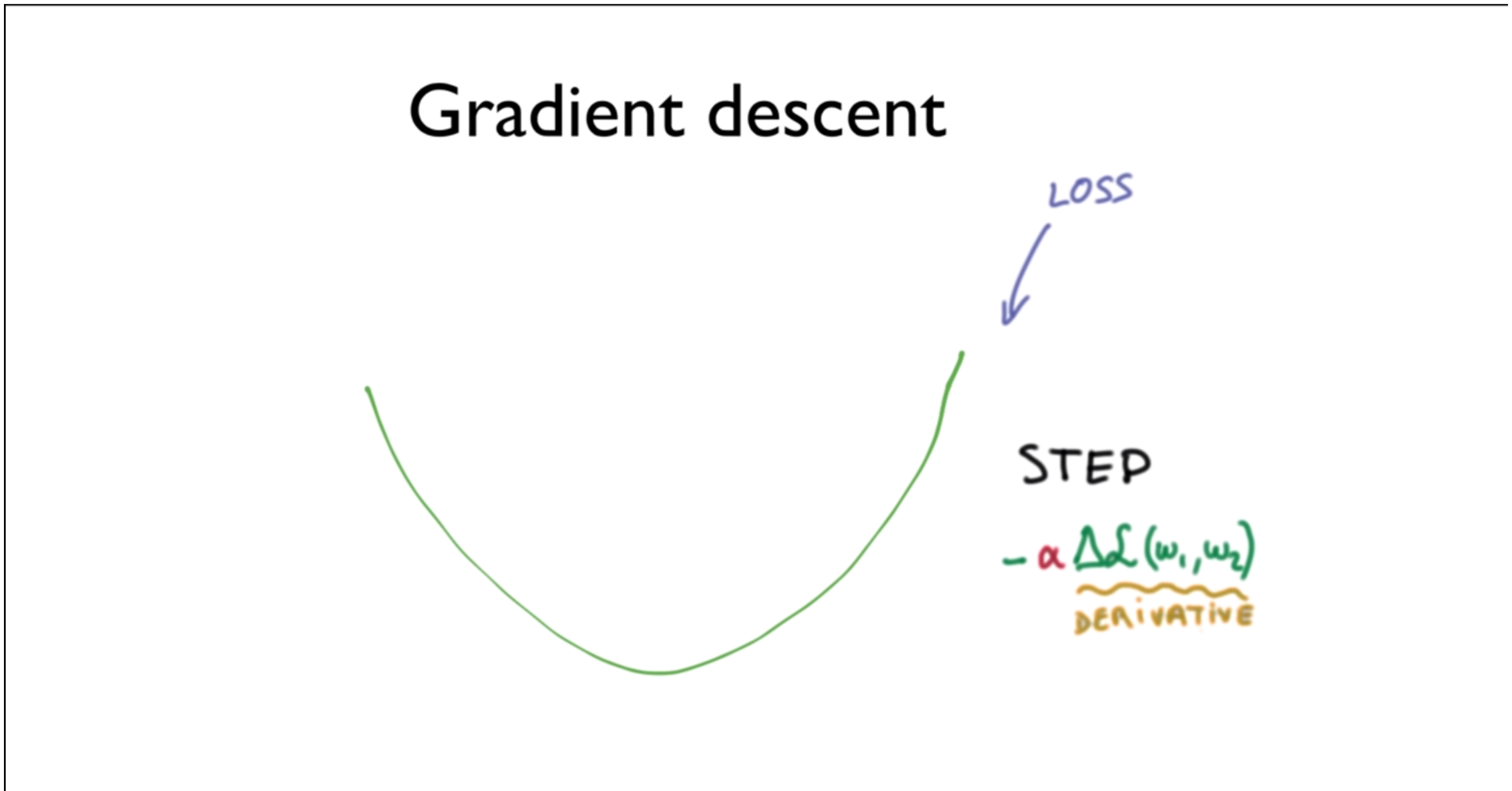
Lecture7

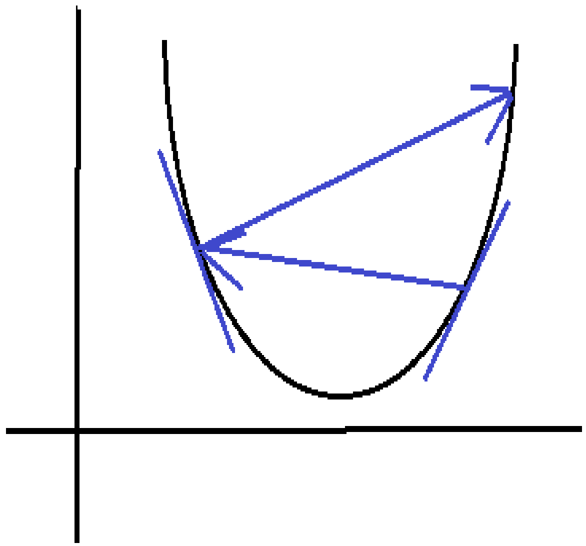
- Learning Rate
- Preprocessing
- Overfitting (Regularization)
- Training/Testing Data Set

Learning Rate



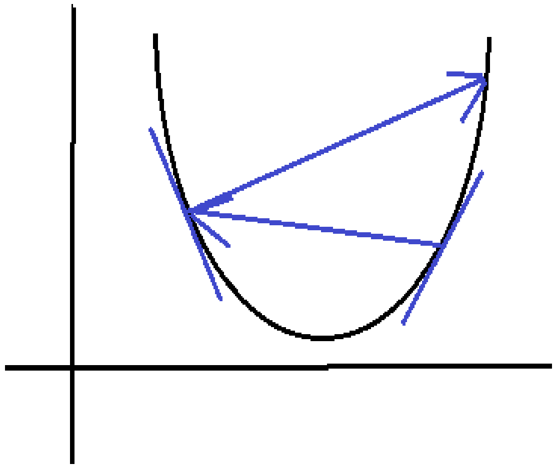
https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811827

Learning Rate

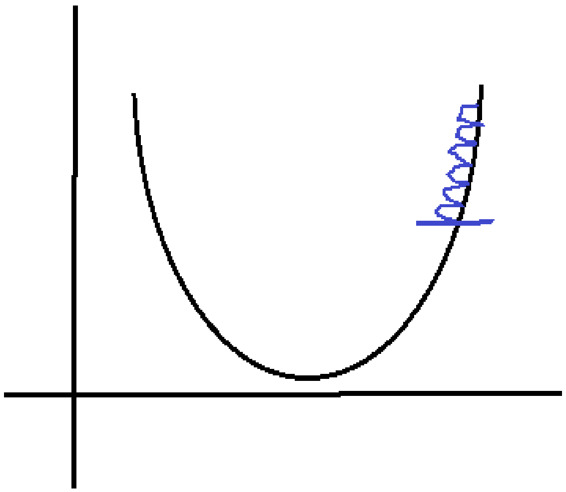


Learning Rate

Lr 너무 클 때

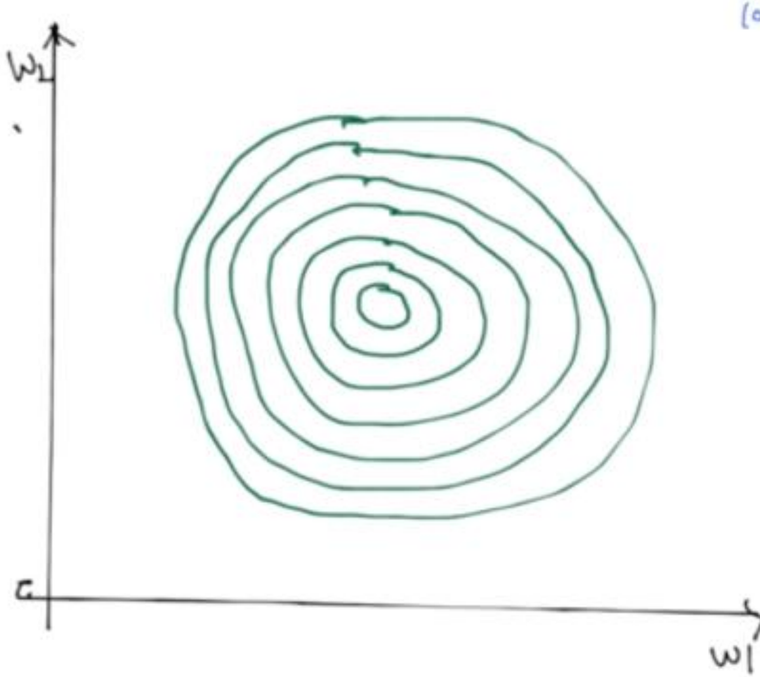


Lr 너무 작을 때



Preprocessing

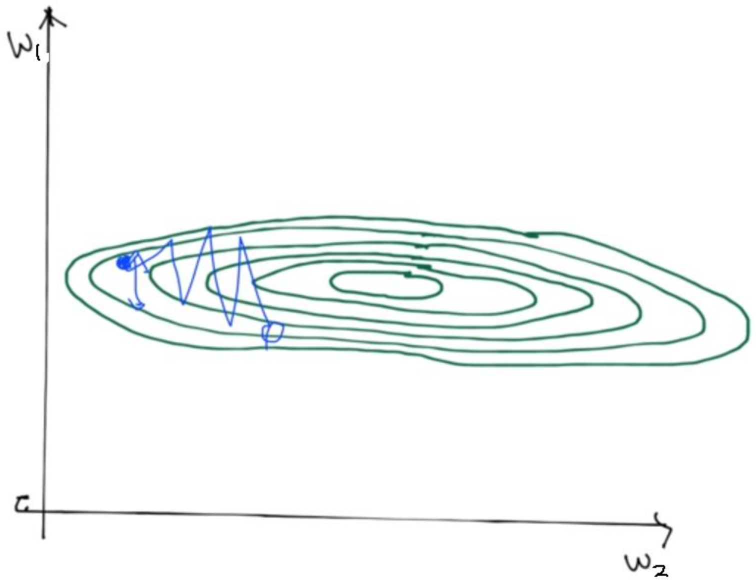
Data (X) preprocessing for gradient descent



Preprocessing

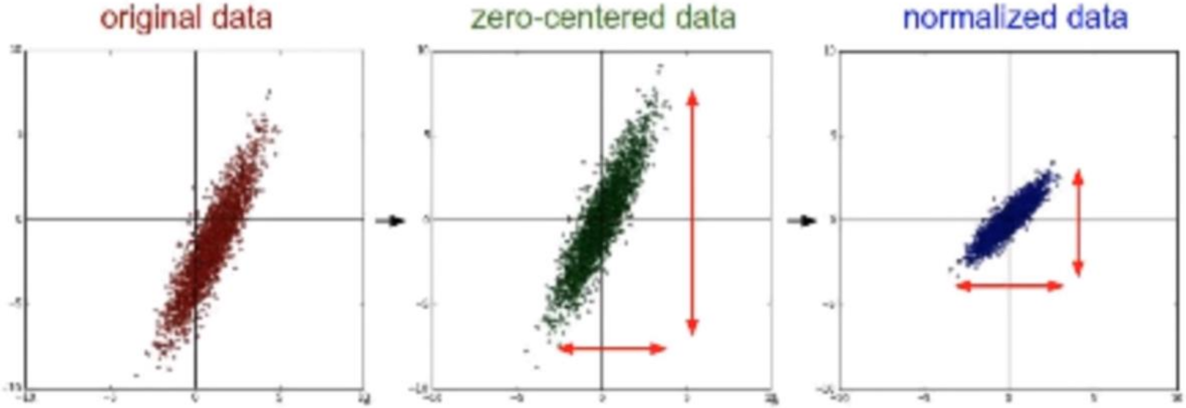
Data (X) preprocessing for gradient descent

x1	x2	y
1	9000	A
2	-5000	A
4	-2000	B
6	8000	B
9	9000	C



Preprocessing

Data (X) preprocessing for gradient descent



Preprocessing

2. 정규화(normalization)

정규화는 데이터의 범위를 0과 1로 변환하여 데이터 분포를 조정하는 방법이다.

(해당 값- 최소값) / (최대값-최소값) 을 해주면 된다.

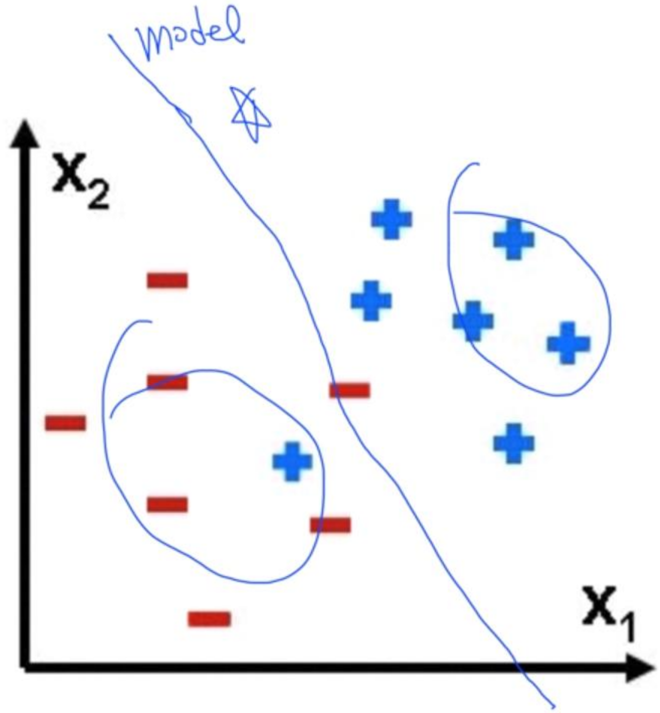
$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

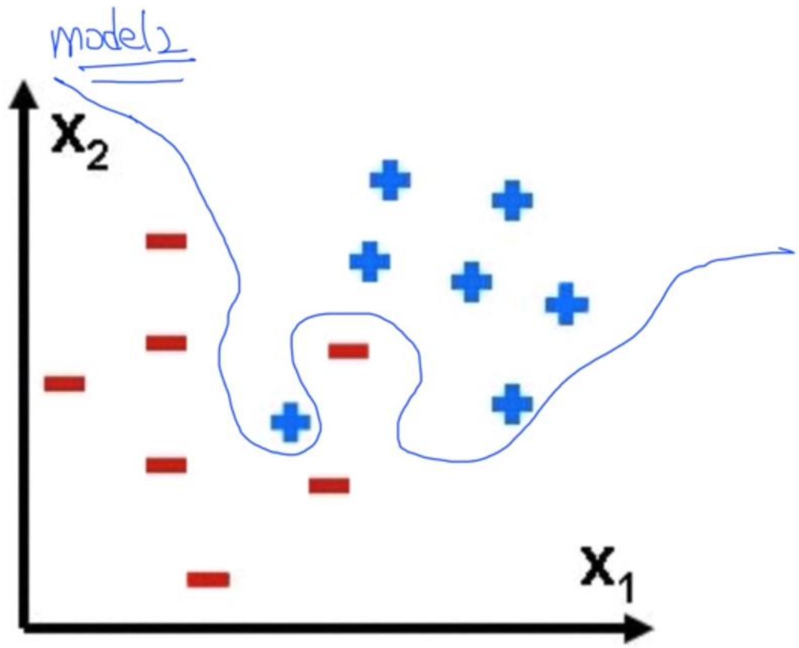
1. 표준화(standardization)

각 observation이 평균을 기준으로 어느 정도 떨어져 있는지를 나타낼때 사용된다. 값의 스케일이 다른 두 개의 변수가 있을 때, 이 변수들의 스케일 차이를 제거해 주는 효과가 있다. 제로 평균 으로 부터 각 값들의 분산을 나타낸다. 각 요소의 값에서 평균을 뺀 다음 표준편차로 나누어 준다.

$$x_{new} = \frac{x - \mu}{\sigma}$$

Overfitting





Overfitting

- Solutions

  More training data

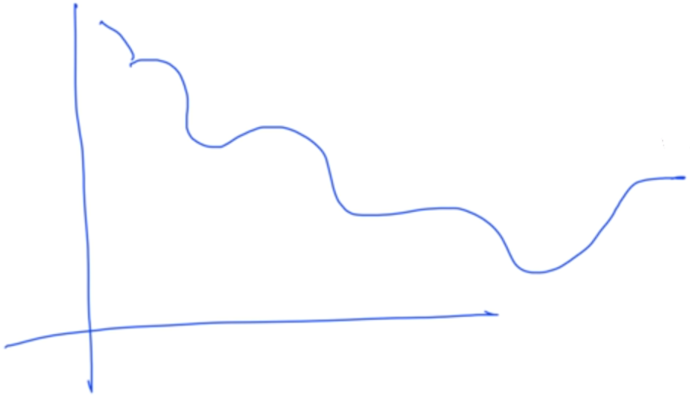
  Reduce the number of features

  Regularization

Overfitting

Regularization

- Let's not have too big numbers in the weight



Overfitting

- Let's not have too big numbers in the weight

LOSS

regularization strength

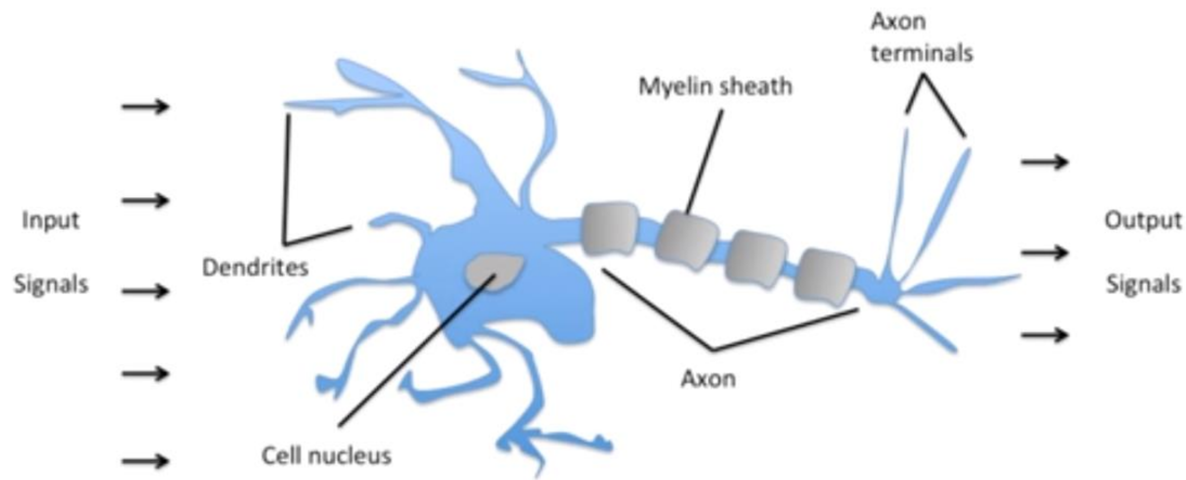
$$\mathcal{L} = \frac{1}{N}\sum_{i} D(S(wx_i+b), L_i) + \lambda \sum W^2$$

TRAINING SET

Lecture 8

- 딥러닝의 시작과 역사

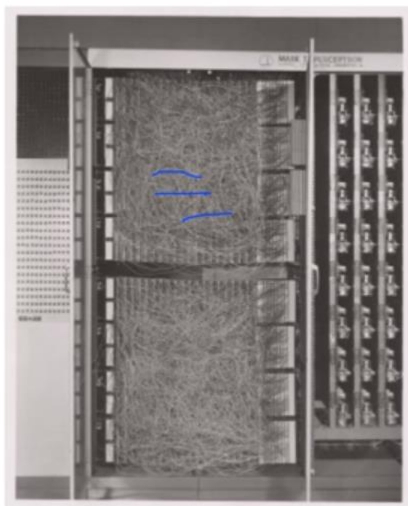
Ultimate dream: thinking machine



Schematic of a biological neuron.

http://sebastianraschka.com/

Hardware implementations



Frank Rosenblatt, ~1957: Perceptron

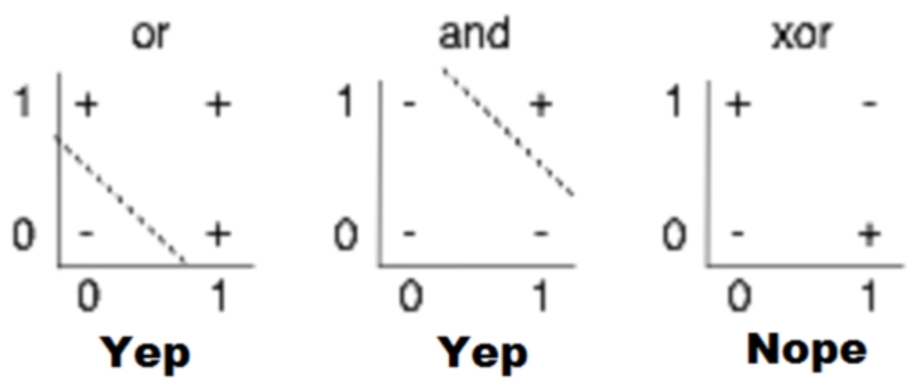
Widrow and Hoff, ~1960: Adaline/Madaline

False Promises

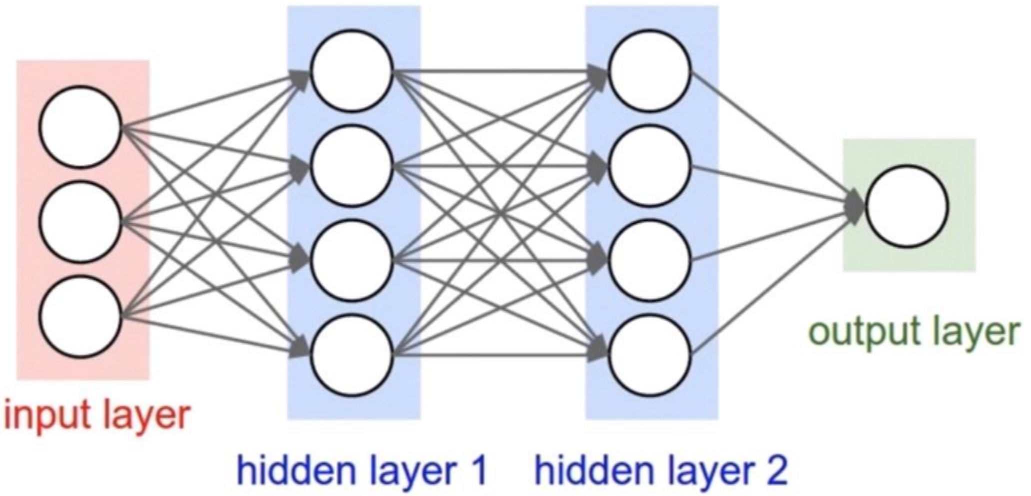
"The Navy revealed the embryo of an electronic computer today that *it expects will be able to walk, talk, see, write, reproduce itself* an be conscious of its existence … Dr. Frank Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers" The New York Times July 08, 1958

http://query.nytimes.com/gst/abstract.html?res=9D01E4D8173DE53BBC4053DFB1668383649EDE

(Simple) XOR problem: linearly separable?



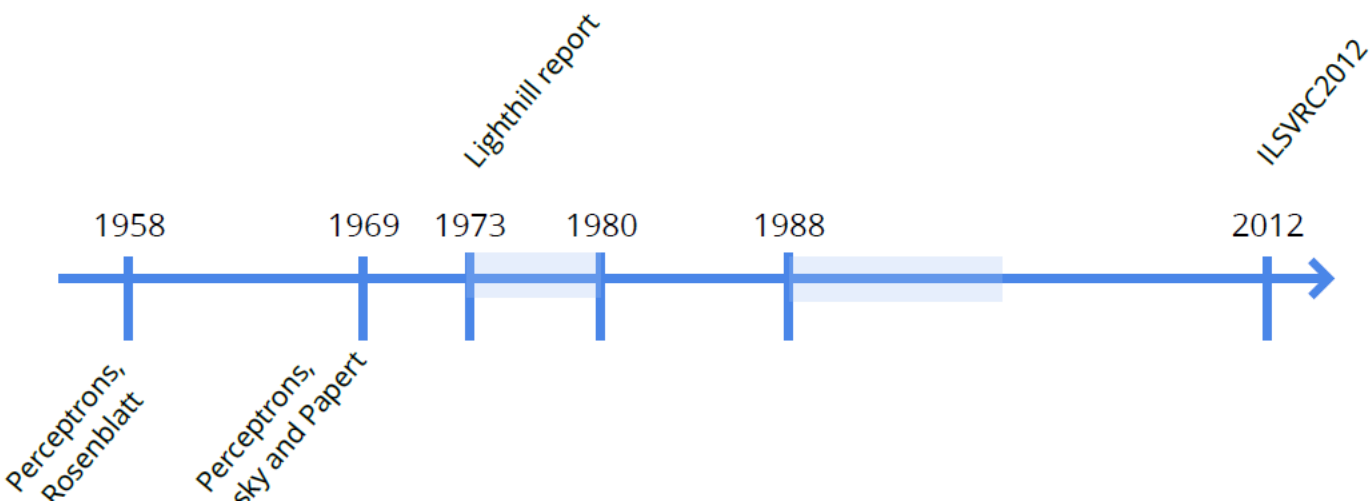
“No one on earth had found a viable way to train*”



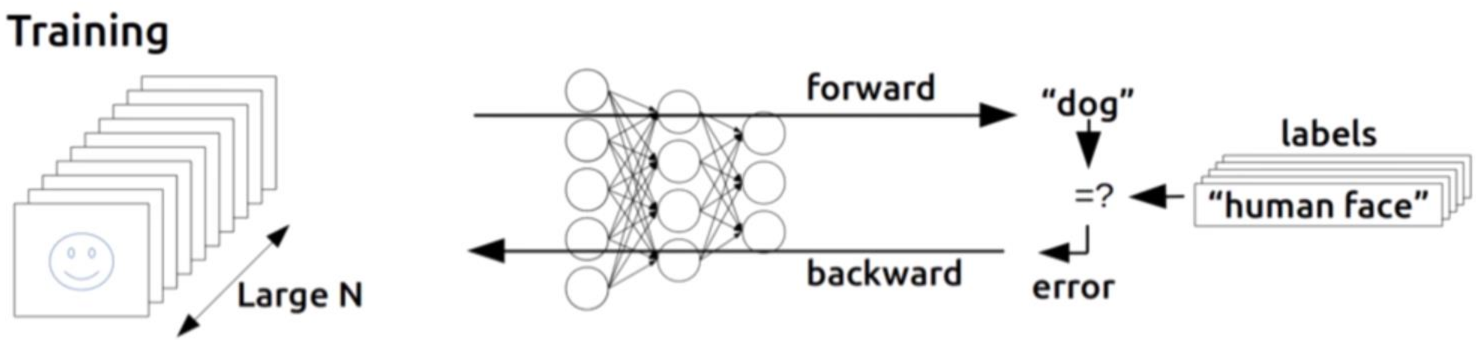
*Marvin Minsky, 1969

http://cs231n.github.io/convolutional-networks/

The AI Winter

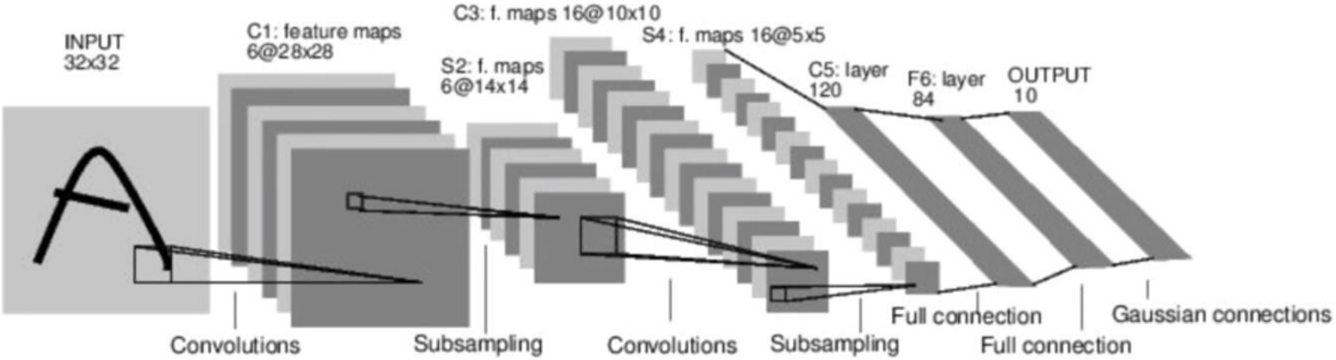


Backpropagation
(1974, 1982 by Paul Werbos, 1986 by Hinton)



https://devblogs.nvidia.com/parallelforall/inference-next-step-gpu-accelerated-deep-learning/

Convolutional Neural Networks

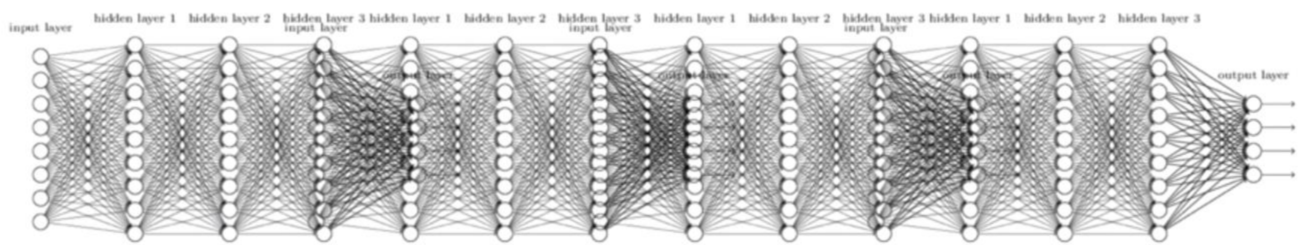


"At some point in the late 1990s, one of these systems was reading 10 to 20% of all the checks in the US."

[LeNet-5, LeCun 1980]

A BIG problem

- Backpropagation just did not work well for normal neural nets with many layers
- Other rising machine learning algorithms: SVM, RandomForest, etc.
- **1995** “Comparison of Learning Algorithms For Handwritten Digit Recognition” by LeCun et al. found that this new approach worked better



http://neuralnetworksanddeeplearning.com/chap6.html

CIFAR

- Canadian Institute for Advanced Research (CIFAR)
- CIFAR encourages basic research *without direct application*, was what motivated **Hinton** to move to Canada in 1987, and funded his work afterward.

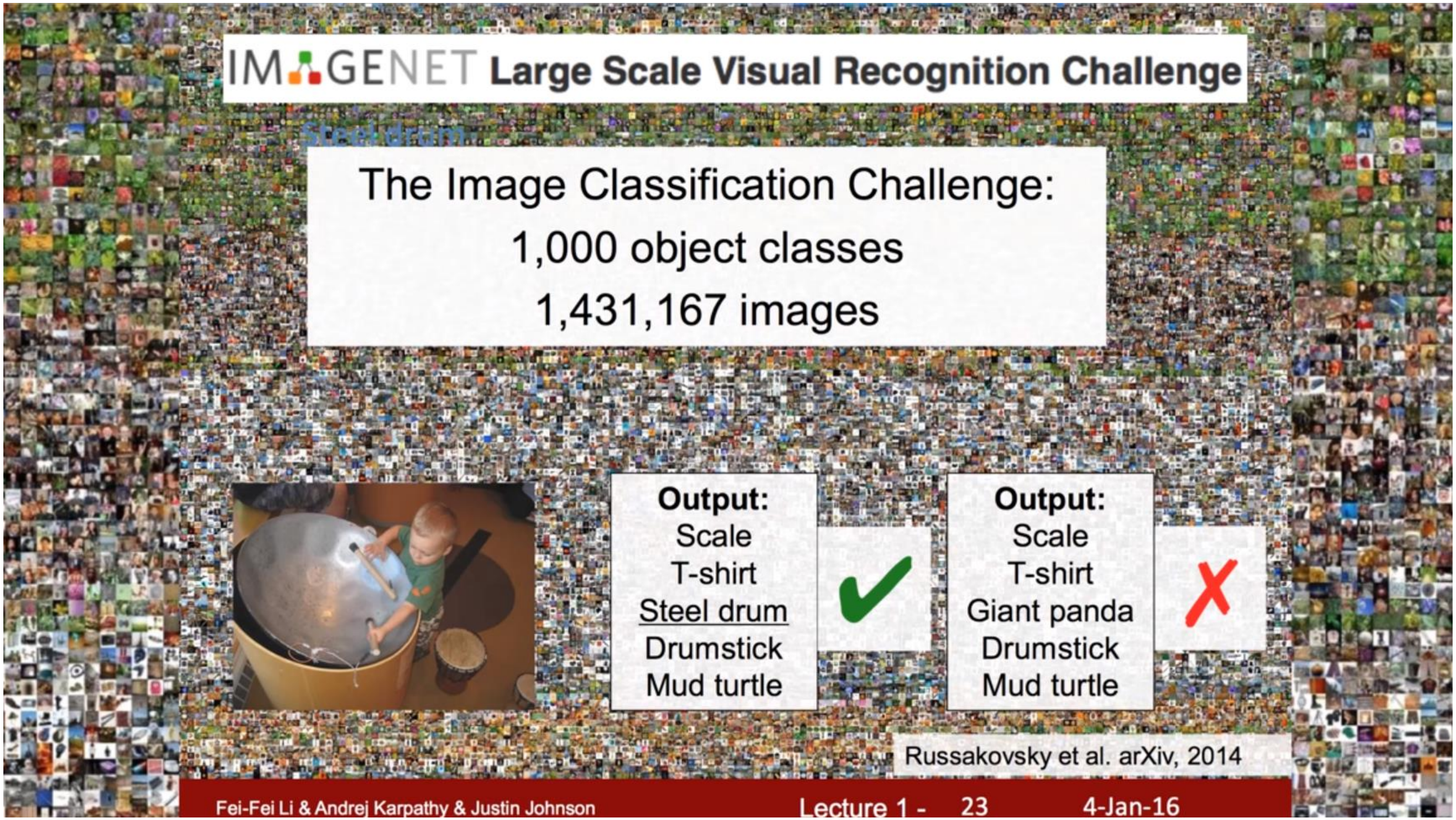
http://www.andreykurenkov.com/writing/a-brief-history-of-neural-nets-and-deep-learning-part-4/

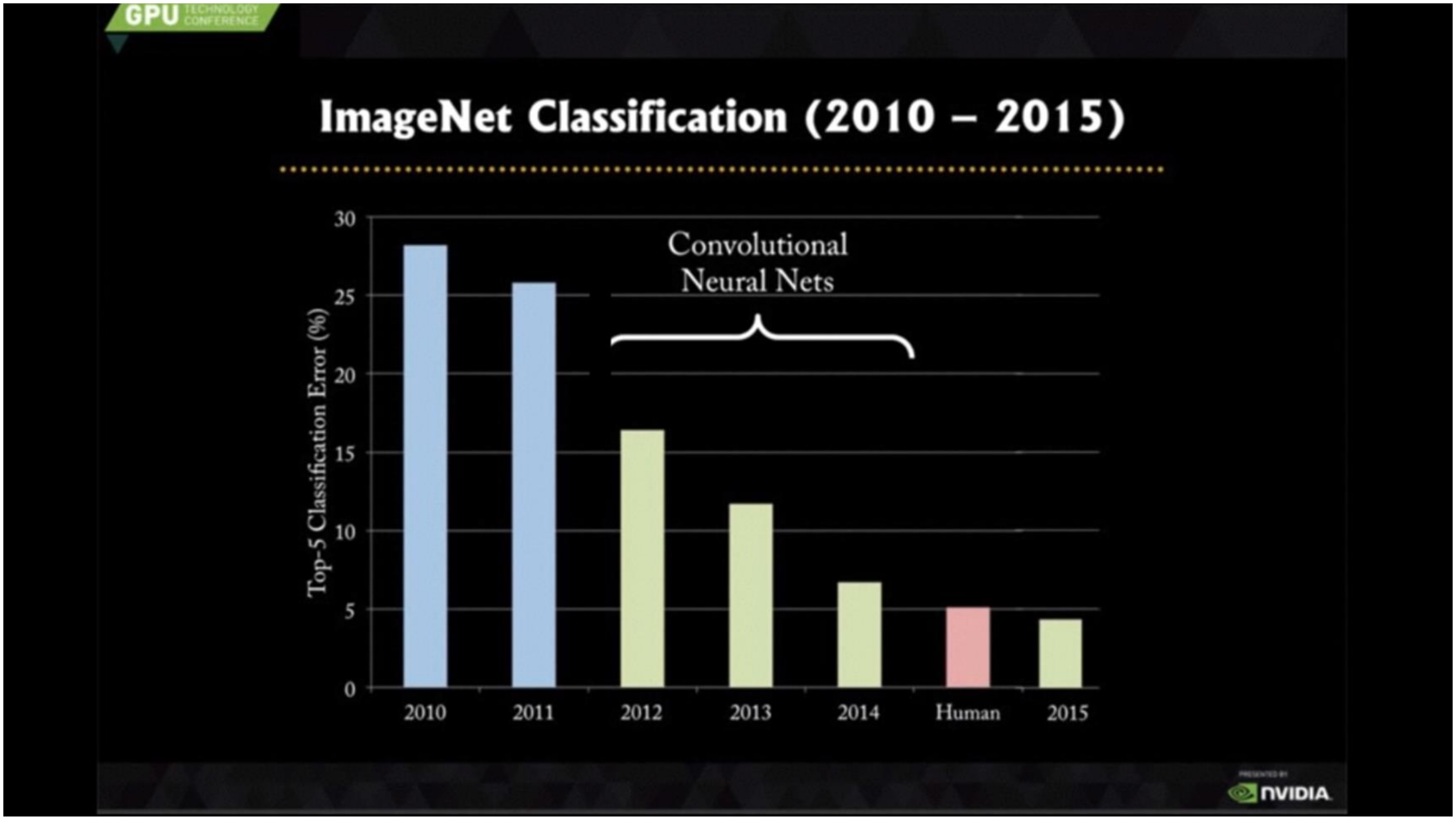
Breakthrough

in 2006 and 2007 by Hinton and Bengio

- Neural networks with many layers really could be trained well, *if the weights are initialized in a clever way* rather than randomly.
- Deep machine learning methods are more efficient for difficult problems than shallow methods.
- Rebranding to ***Deep Nets***, ***Deep Learning***

http://www.andreykurenkov.com/writing/a-brief-history-of-neural-nets-and-deep-learning-part-4/


IMAGENET Large Scale Visual Recognition Challenge
The Image Classification Challenge:
1,000 object classes
1,431,167 images
Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle
Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle
Russakovsky et al. arXiv, 2014
Fei-Fei Li & Andrej Karpathy & Justin Johnson
Lecture 1 - 23
4-Jan-16


GPU TECHNOLOGY CONFERENCE
ImageNet Classification (2010 – 2015)
Convolutional
Neural Nets
Top-5 Classification Error (%)
30
25
20
15
10
5
0
2010
2011
2012
2013
2014
Human
2015
NVIDIA


ALPHAGO
01:58:30
LEE SEDOL
Mind
Match
AlphaGo
Lee Sedol