機器學習現在、過去、未來
A Little History of Machine Learning

Chia-Ping Chen

National Sun Yat-sen University

@NPTU, December 2016

Outline

- ubiquitous machine intelligence
- challenge and reaction
- AI brief
- deep learning
- conclusion

ubiquitous machine intelligence

San Francisco: 673 MUNI

Seoul: Traveling Salesman

Just Very Recently

A trip to Bali

- **advertisement**
- **map**
- **immigration**

Daily life

- **bus/train/THSR**
- **i-card**
- **bridge**

---

information system *vs.* **intelligent system**

big data *vs.* **machine learning**

The Largest Slice of Pie

device (hardware + software)

networking

machine learning

Who Wants to Do Machine Learning

- web giants
- device manufacturers
- internet service providers
- social networks
- e-commerce
- chip design/manufacture company

Google + Facebook + Amazon + IBM + MS

challenge and reaction

Machines Today

Robots

- cooking, making pizza
- playing piano, trumpet
- listening and talking
- driving
- playing games

Provide (replace) manpowers in

- stock market
- medicine
- law
- education

47% 77%

Who Is Better? Human or Machine

- speaking
- painting
- singing
- writing
- games
- question answering
- special effects

Human in Machine Era

- recognize and acknowledge the intrinsic difference between human and machine
- use machines
- learn machine learning
- deep better than shallow
- creative better than routine
- do what you like and like what you do
- humanity
- heart
- emotion

AI brief

Legends

Church-Turing thesis

"A function on the natural numbers is computable by a human being following an algorithm, ignoring resource limitations, if and only if it is computable by a Turing machine."

Turing test

"A human being should be unable to distinguish the machine from another human being by using the replies to questions put to both."

Data Mining

Pattern Recognition

Human PR

Brahe	$\rightarrow$	**Kepler**	$\rightarrow$	**Newton**
(observation)		(description)		(explanation)

Machine PR

computer	$\rightarrow$	**computer**	$\rightarrow$	**who cares?**
(transactions)		(association)		(hindsight)

Machine Learning

"A computer program is said to learn from **experience** E *with respect to some class of* **tasks** T *and* **performance measure** P*, if its performance at tasks in* T*, as measured by* P*, improves with experience* E*."*

Tasks and Performances

- **anomaly detection**
- **classification**
- **regression**
- **transcription**
- **translation**
- **synthesis**
- parsing
- imputation
- denoising
- density estimation

deep learning

Milestones

- Deep Blue
- Deep QA
- AlphaGo (by DeepMind)

Going Deep

- VoiceSearch (automatic speech recognition)
- WaveNet (speech synthesis)
- ImageNet (image classification)
- Translator (machine translation)
- AlphaGo (honorary 10 dan)

LeCun, Bengio, and Hinton, "Deep Learning", Nature, 2015

D. Silver, Aja Huang, et al. "Mastering the game of Go with deep neural networks and tree search", Nature, 2016

Future of Machine Learning

- robots (Tokyo 2020)
- imitating human
- creativity (synthesis of art works)
- unsupervised learning
- zero-data learning
- reinforcement learning
- long-term and long-range problems (programming, writing)
- going after really difficult problems (weather, earthquake, astronomy, brain, cancer, medicine)

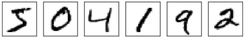
a research work at NSYSU MIT Lab

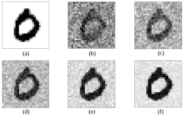
Background

- accepted for oral presentation at **ICASSP 2016**
- joint work with **Hon Hai Technology Group**
- simple idea, good results
- sleepless in San Franscisco

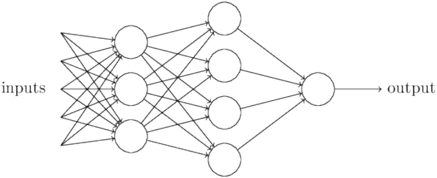
Problem Statement

- **task:** hand-written digit recognition
- **data:** images of hand-written digits with added noises





- **goal:** noise-robustness
- **model:** artificial neural network



Artificial Neural Network

- **input layer** for the pixels

$$\mathbf{x} = \{x_i\}$$

- **hidden layer** to bridge between input and output

$$\mathbf{h} = \{h_k\}$$

- **output layer** for the digits

$$\mathbf{y} = \{y_j\}$$

Propagation of Information

- linear combination of input pixel values

$$a_k = \sum_{m=1}^{M} \sum_{n=1}^{N} w_{k,mn} x_{mn}$$

  where x_{mn} is the pixel value at position (m, n), and $w_{k,mn}$ is the **weight** of the link from position (m, n) to unit k
- non-linear **activation function**

$$h_k = \sigma(a_k)$$

- linear combination of hidden unit values

$$a_j = \sum_{k=1}^{K} w_{jk} h_k$$

- transform a_j to output value y_j

Vector vs. Image

An image can be represented by a vector, and vice versa.

- image to vector

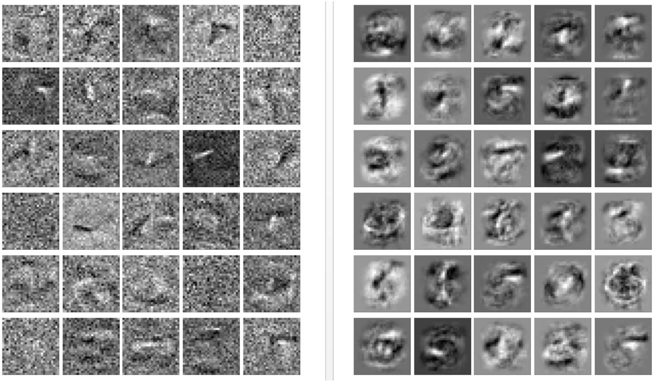
$$\begin{bmatrix} x_{11} & \dots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{M1} & \dots & x_{MN} \end{bmatrix} \quad \longrightarrow \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_I \end{pmatrix}$$

- vector to image

$$\begin{pmatrix} w_{k,11} \\ \vdots \\ w_{k,MN} \end{pmatrix} = \mathbf{w}_k \quad \longrightarrow \quad \begin{bmatrix} w_{k,11} & \dots & w_{k,1N} \\ \vdots & \ddots & \vdots \\ w_{k,M1} & \dots & w_{k,MN} \end{bmatrix}$$

Visualization

Visualize a hidden unit k by incoming link weights $w_{k,mn}$.



30 images, each corresponds to a hidden unit.
left block: without orthogonalization.
right block: with orthogonalization.

Inner Product and Projection

The linear combination

$$a_k = \sum_{i=1}^{I} w_{ki} x_i$$

can be seen as an **inner product**

$$a_k = \mathbf{w}_k^T \mathbf{x}$$

or as a **projection**, noting that the projection of $\mathbf{x}$ to $\mathbf{v}$ is

$$\frac{\mathbf{v}^T \mathbf{x}}{\mathbf{v}^T \mathbf{v}} \mathbf{v}$$

Orthogonality

We often want to work with an orthogonal basis.

- orthogonality makes projection simple
- orthogonality makes projection economic
- orthogonality makes projection independent

Gram-Schmidt Process

Gram-Schmidt process converts a set of linearly independent vectors

$$\mathbf{v}_1, \ldots, \mathbf{v}_J$$

to a set of **orthonormal** vectors

$$\mathbf{w}_1, \ldots, \mathbf{w}_J$$

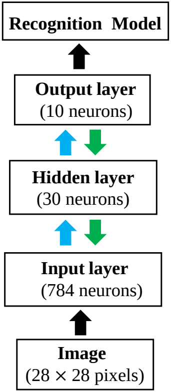
by **orthogonalization**

$$\mathbf{u}_j = \mathbf{v}_j - \sum_{i=1}^{j-1} \langle \mathbf{v}_j, \mathbf{u}_i \rangle \mathbf{u}_i$$

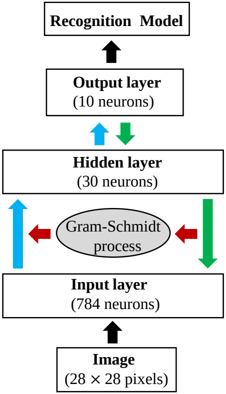
and **normalization**

$$\mathbf{w}_j = \frac{\mathbf{u}_j}{|\mathbf{u}_j|}$$

System Framework



baseline



proposed

MNIST

- images of hand-written digits, each with 28×28 pixels

$$I = 784$$

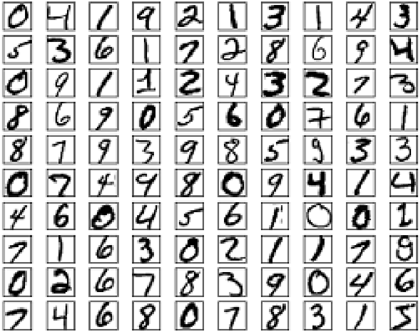
- 10 classes for digit 0 to digit 9

$$J = 10$$

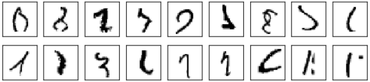
- 50000 training examples and 10000 test examples

Examples

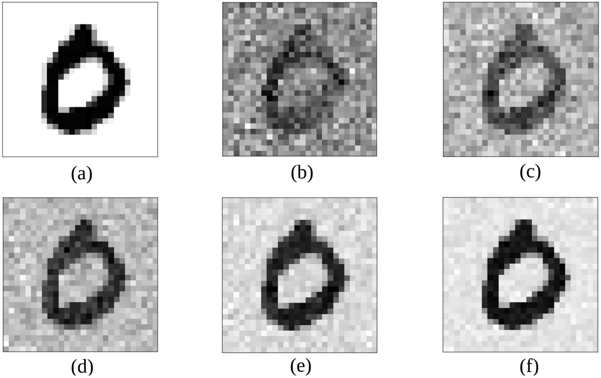
random



difficult

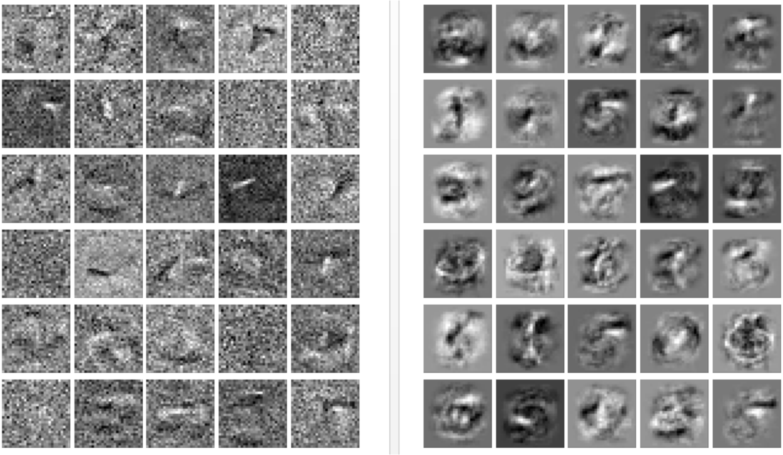


Noisy Data



clean and noisy data with 0-20 dB SNR

Weight Images At a Glance

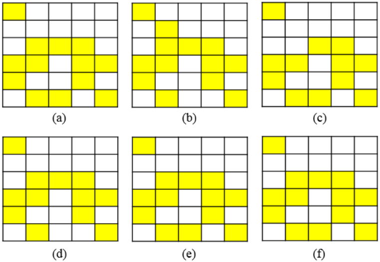


Results

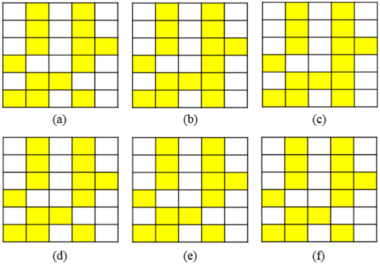
baseline		
SNR	accuracy	imp. over clean
clean	95.1	=
20 dB	94.9	-4.1
15 dB	94.5	-12.2
10 dB	93.6	-30.1
5 dB	91.4	-75.5
0 dB	81.9	-269.4

proposed		
SNR	accuracy	imp. over baseline
20 dB	94.3	-12.2
15 dB	95.0	9.1
10 dB	95.3	26.6
5 dB	94.6	38.1
0 dB	**92.1**	**56.4**

Visualization of Hidden Layer



$$\mathbf{h}_1, \ldots, \mathbf{h}_6, \quad \mathbf{h}_i = \sigma(\mathbf{a}_i)$$



$$\mathbf{h}'_1, \ldots, \mathbf{h}'_6, \quad \mathbf{h}'_i = \sigma(\mathbf{a}'_i)$$

Hamming distance between clean and noisy data

- baseline system

$$2, 1, 1, 0, 0$$

- proposed system

$$1, 1, 0, 0, 0$$

Conclusion

- machine learning in daily lives
- brief AI history
- booming deep learning
- machine learning research @ NSYSU MIT Lab
    - EmoDB and FAU-Aibo
    - MNIST
    - Aurora
    - Twitters
    - Board games