

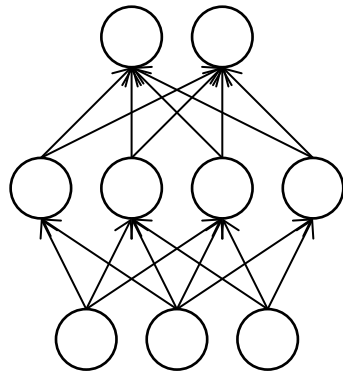
ディープラーニング

チュートリアル（もしくは研究動向報告）

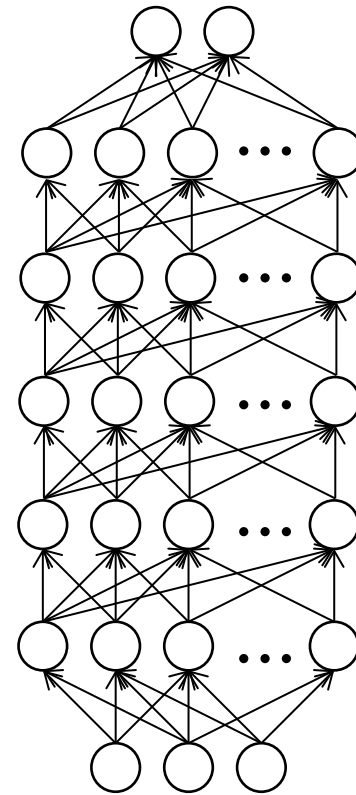
岡谷 貴之 齋藤 真樹
東北大学

Deep Learningとは

- Deep Learning = 多層ニューラルネットを使った機械学習の方法論
- ニューラルネットの「ルネッサンス」



Shallow NN



Deep Neural Network (DNN)

目次

- インTRODクシヨN
 - DNNの事例紹介
- NNの基礎と研究の略史
 - なぜNN研究は廃れたか？
- プレトレーニング
 - Encoder-decoder paradigm
- Convolutional Neural Networks (CNN)
 - プーリングと不変特徴
 - CNN vs. fully-connected NN
- ニューロサイエンス
 - どこまで分かっている？
- 生成モデル
 - Restricted Boltzmann Machine (RBM)
 - Deep Belief Network (DBN)
- 実践編
 - cuda-convnet を使ったMNISTの学習
- まとめ

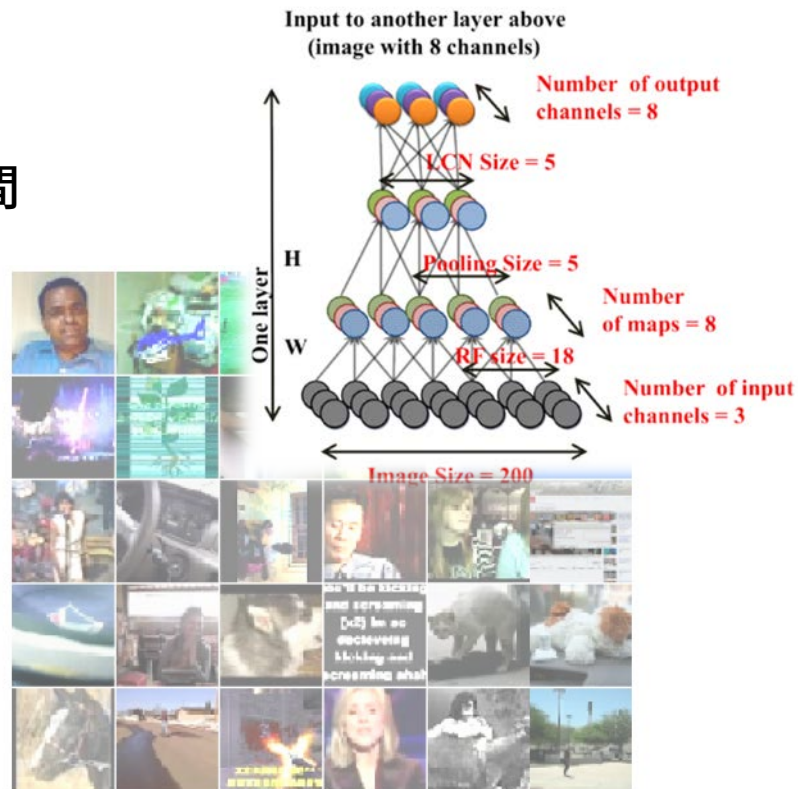
目次

- **イントロダクション**
 - **DNNの事例紹介**
- **NNの基礎と研究の略史**
 - なぜNN研究は廃れたか？
- **プレトレーニング**
 - Encoder-decoder paradigm
- **Convolutional Neural Networks (CNN)**
 - プーリングと不変特徴
 - CNN vs. fully-connected NN
- **ニューロサイエンス**
 - どこまで分かっている？
- **生成モデル**
 - Restricted Boltzmann Machine (RBM)
 - Deep Belief Network (DBN)
- **実践編**
 - cuda-convnet を使ったMNISTの学習
- **まとめ**

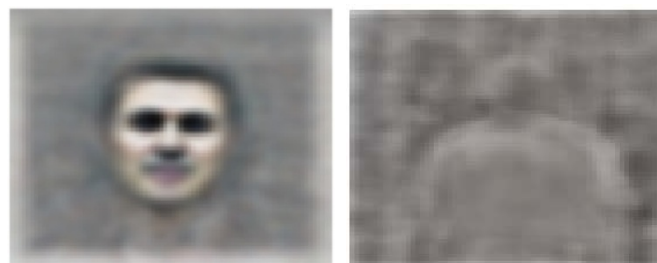
事例 1 : 画像特徴の無教師学習 (Ng のグループ)

Le et al., Building High-level Features Using Large Scale Unsupervised Learning, ICML2012

- 12層NNを使った無教師学習
 - パラメータ数10億個!
 - 16コアPC1000台のPCクラスター×3日間
 - YouTubeの画像1000万枚
- 「おばあさん細胞」の生成を確認



The New York Times (2012/6/25)



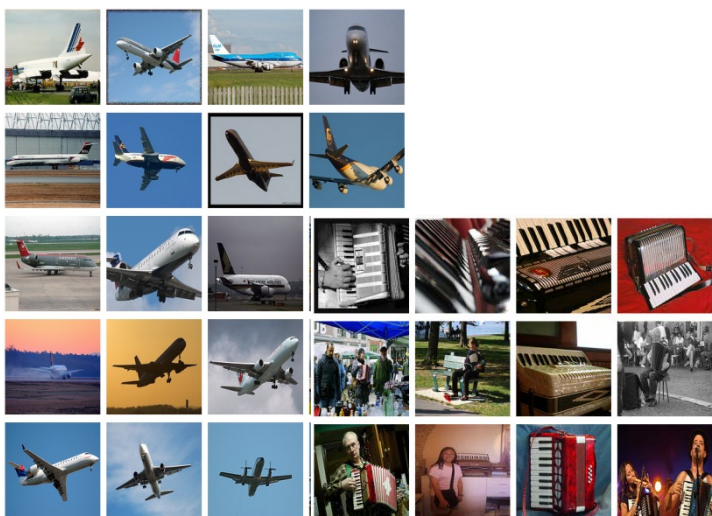
顔

人の体

事例 2: 一般物体認識 (Hintonのグループ)

Krizhevsk et al., ImageNet Classification with Deep Convolutional Neural Networks, NIPS2012

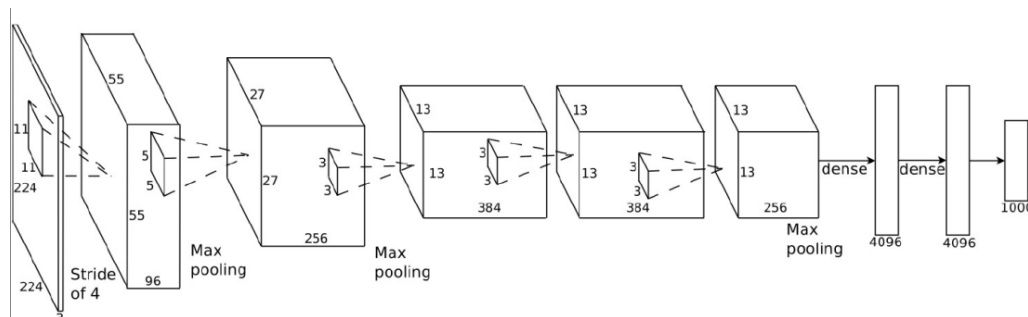
- IMAGENET Large Scale Visual Recognition Challenge 2012
 - 1000カテゴリ・カテゴリあたり約1000枚の訓練画像



“airliner”

“accordion ...”

	Team name	Error (5 guesses)
1	SuperVision	0.15315
2	ISI	0.26172
3	OXFORD_VGG	0.26979
4	XRCE/INRIA	0.27058
5	University of Amsterdam	0.29576
6	LEAR-XRCE	0.34464



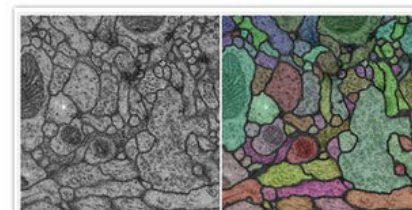
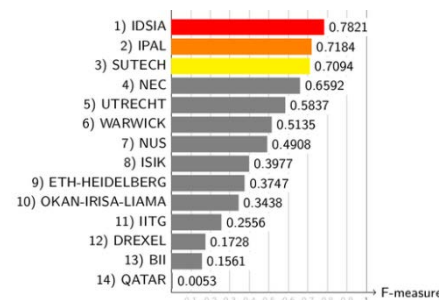
7-layer NN

事例 3 : 文字・画像認識 (Schmidhuberのグループ@IDSIA)



• コンテスト1位

- IJCNN 2011 Traffic Sign Recognition Competition; 1st (0.56%), 2nd (1.16%, Humans), 3rd (1.69%), 4th (3.86%)
- ICPR 2012 Contest on "Mitosis Detection in Breast Cancer Histological Images"
- ISBI 2012 challenge on segmentation of neuronal structures
- ICDAR 2011 Offline Chinese Handwriting Competition
- ICDAR2009
 - Arabic Connected Handwriting Competition
 - French Connected Handwriting Competition



• 認識率最高位 (2012年11月時点)

- NORB object recognition benchmark
- CIFAR image classification benchmark
- MNIST handwritten digits benchmark; "human-competitive result"

Dataset	Best result of others [%]	MCDNN [%]	Relative improv. [%]
MNIST	0.39	0.23	41
NIST SD 19	see Table 4	see Table 4	30-80
HWDB 1.0 on.	7.61	5.61	26
HWDB 1.0 off.	10.01	6.5	35
CIFAR10	18.50	11.21	39
traffic signs	1.69	0.54	72
NORB	5.00	2.70	46

事例：音声認識 (speech recognition)

Hinton et al., Deep Neural Networks for Acoustic Modeling in Speech Recognition, IEEE SP magazine, Nov. 2012

- 問題
 - 入力：音声信号から取り出したMFCC (Mel-frequency cepstral coefficients)
 - 出力：HMMのステート
- 過去のstate-of-the-art: GMM-HMM
- 入出力関係をDNNで学習 → **30%の性能向上**

[TABLE 3] A COMPARISON OF THE PERCENTAGE WERs USING DNN-HMMs AND GMM-HMMs ON FIVE DIFFERENT LARGE VOCABULARY TASKS.

TASK	HOURS OF TRAINING DATA	DNN-HMM	GMM-HMM WITH SAME DATA	GMM-HMM WITH MORE DATA
SWITCHBOARD (TEST SET 1)	309	18.5	27.4	18.6 (2,000 H)
SWITCHBOARD (TEST SET 2)	309	16.1	23.6	17.1 (2,000 H)
ENGLISH BROADCAST NEWS	50	17.5	18.8	
BING VOICE SEARCH (SENTENCE ERROR RATES)	24	30.4	36.2	
GOOGLE VOICE INPUT	5,870	12.3		16.0 (>> 5,870 H)
YOUTUBE	1,400	47.6	52.3	

事例：化合物反応予測

Merck Molecular Activity Challenge

Finished

Thursday, August 16, 2012

\$40,000 • 238 teams Tuesday, October 16, 2012

Competition Details » Get the Data » Make a submission

Help develop safe and effective medicines by predicting molecular activity.

Help enable the development of safe, effective medicines.

When **developing new medicines** it is important to identify molecules that are highly active toward their intended targets but not toward other targets that might cause side effects. The objective of this competition is to identify the best statistical techniques for predicting biological activities of different molecules, both on- and off-target, given numerical descriptors generated from their chemical structures.

The challenge is based on 15 molecular activity data sets, each for a biologically relevant target. Each row corresponds to a molecule and contains descriptors derived from that molecule's chemical structure.

In addition to the prediction competition, Merck is also hosting a **visualization challenge** with a \$2,000 prize for the most insightful and elegant graphical representation of the data.

Prizes total **\$40,000**.

Started: 9:34 pm, Thursday 16
Ended: 11:59 pm, Tuesday 16 October

- Description
- Evaluation
- Rules
- Prizes
- Submission Instructions
- Visualization Prospect
- Winners

Winners

Read the Winner's [Announcement](#) on No Free Hunch.

First Place

- George Dahl, Toronto, Canada
- Ruslan Salakhutdinov, Toronto, Canada
- Navdeep Jaitly, Toronto, Canada
- Chris Jordan-Squire, Seattle, Washington
- Geoffrey Hinton, Toronto, Canada

Second Place

- Jeremy Achin, Hartford CT, USA
- Tom DeGodoy, Hartford CT, USA
- Xavier Conort, Singapore

Third Place

- Alexander A. Larko, Minusinsk, Russia
- Eu Jin Lok, Melbourne, Australia
- Zach Mayer, Boston MA, USA

Scientists See Promise in Deep-Learning Programs



Hao Zhang/The New York Times

A voice recognition program translated a speech given by Richard F. Rashid, Microsoft's top scientist, into Mandarin Chinese.

By JOHN MARKOFF

Published: November 23, 2012

Using an artificial intelligence technique inspired by theories about how the brain recognizes patterns, technology companies are reporting startling gains in fields as diverse as computer vision, speech recognition and the identification of promising new molecules for designing drugs.

点击查看本文中文版。

Connect With Us on Social Media



@nytimescience on Twitter.

Science Reporters and Editors on Twitter

Like the science desk on Facebook.



Keith Penner

The advances have led to widespread enthusiasm among researchers who design software to perform human activities like seeing, listening and thinking. They offer the promise of machines that converse with humans and perform tasks like driving cars and working in factories, [raising the specter of automated robots that could replace human workers](#).

The technology, called deep learning, has already been put to use in services like Apple's Siri virtual personal assistant, which is based on Nuance Communications' speech recognition service, and in Google's Street View, which uses machine vision to identify specific addresses.

But what is new in recent months is the growing speed and accuracy of deep-learning programs, often called artificial neural networks or just "neural nets" for their resemblance to the neural connections in the brain.

"There has been a number of stunning new results with deep-learning methods," said Yann LeCun, a

- FACEBOOK
- TWITTER
- GOOGLE+
- SAVE
- E-MAIL
- SHARE
- PRINT
- SINGLE PAGE
- REPRINTS

STOKER
COMING SOON

11/21/2012, The New York Times

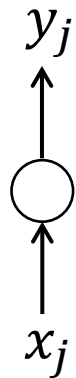
目次

- インTRODクシヨN
 - DNNの事例紹介
- **NNの基礎と研究の略史**
 - **なぜNN研究は廃れたか？**
- プレトレーニング
 - Encoder-decoder paradigm
- Convolutional Neural Networks (CNN)
 - プーリングと不変特徴
 - CNN vs. fully-connected NN
- ニューロサイエンス
 - どこまで分かっている？
- 生成モデル
 - Restricted Boltzmann Machine (RBM)
 - Deep Belief Network (DBN)
- 実践編
 - cuda-convnet を使ったMNISTの学習
- まとめ

NNの基本要素

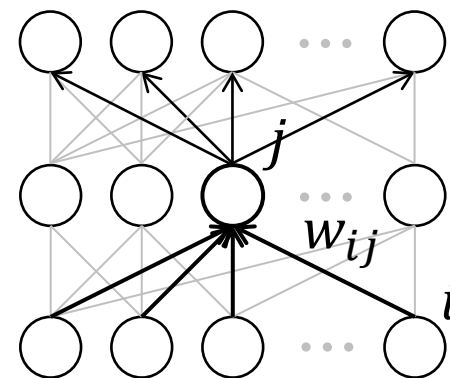
非線形入出力のユニット

ネットワーク



$$y_j = f(x_j)$$

$$x_j = b_j + \sum_i y_i w_{ij}$$



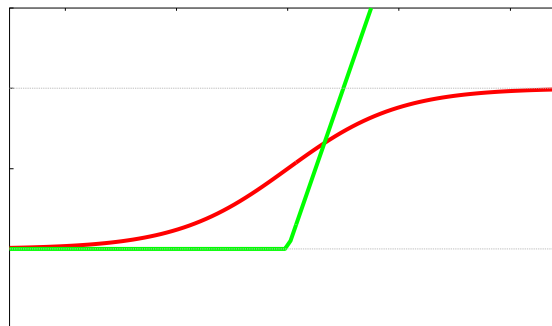
activation function

$$f(x_j) = \frac{1}{1 + e^{-x_j}}$$

$$f(x_j) = \tanh(x_j)$$

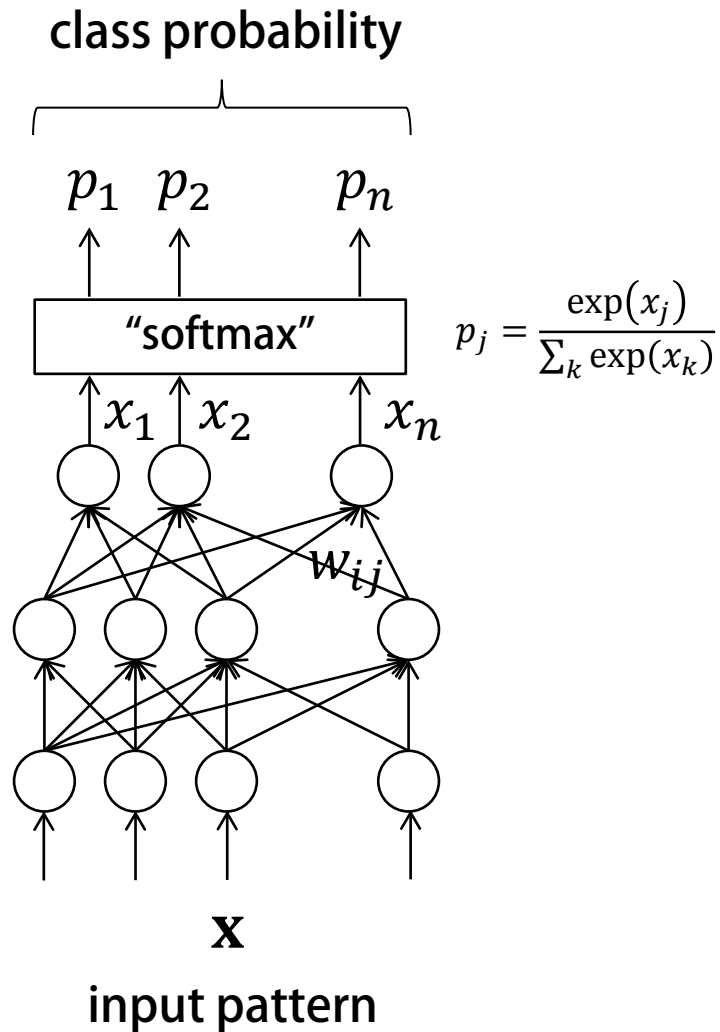
$$f(x_j) = \max(x_j, 0)$$

y_j

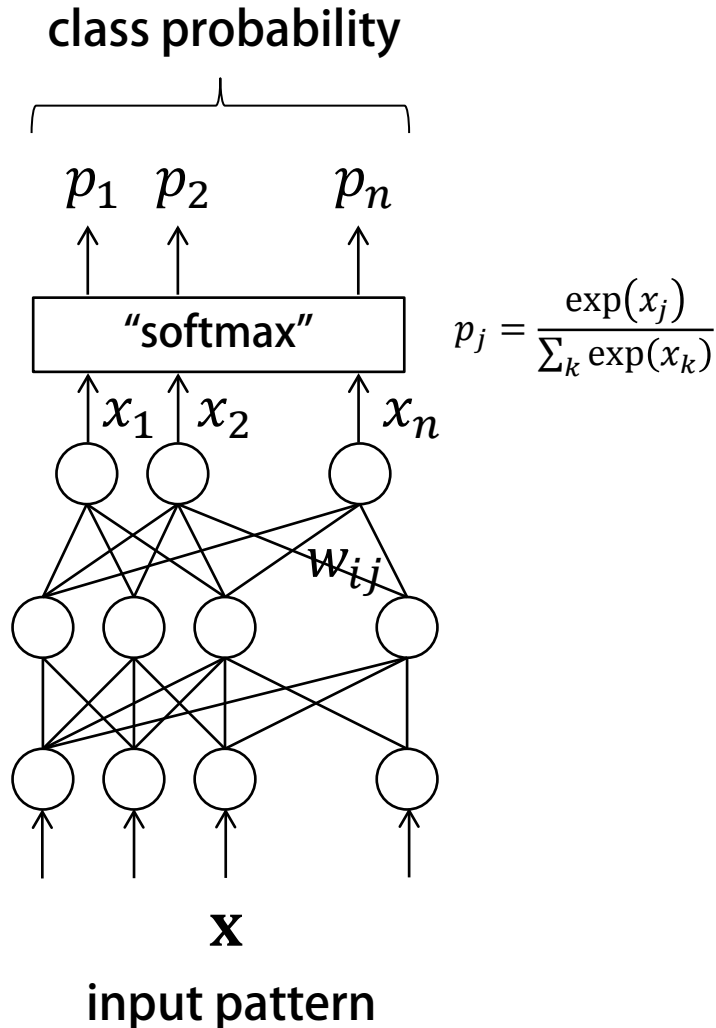


x_j

NNの学習 (分類問題の場合)



NNの学習 (分類問題の場合)



学習サンプル $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$ を再現するようにネットワークの重み $\{w_{ij}\}, \{b_j\}$ を調節

誤差 (=cross entropy) を最小化

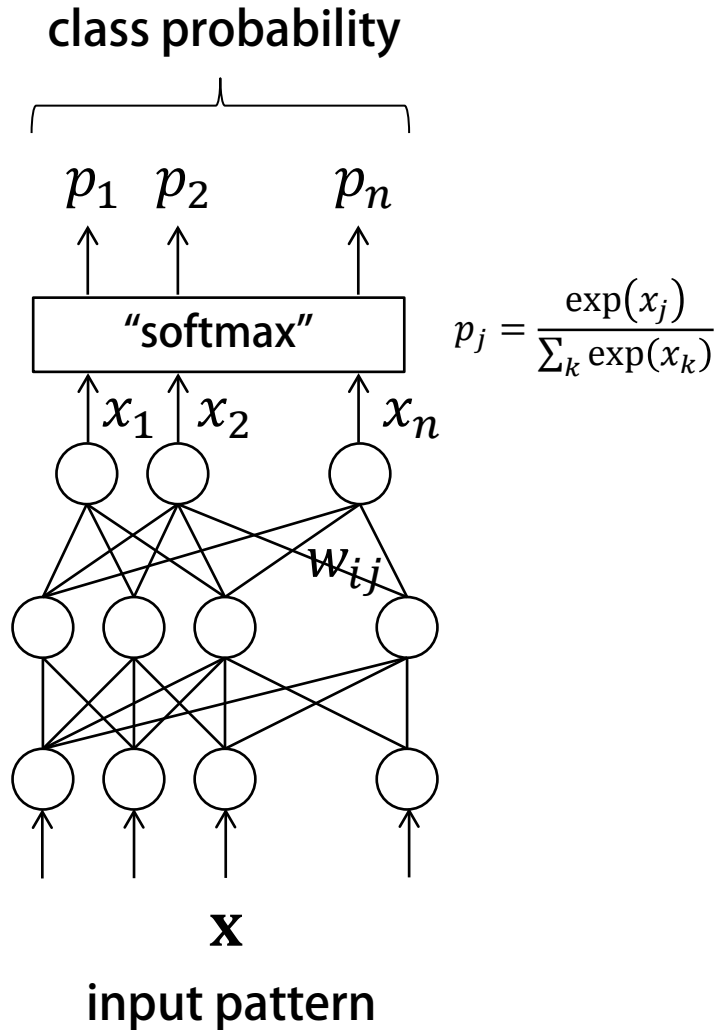
$$C = - \sum_j d_j \log p_j$$

(理想的には全学習サンプルの総和)

勾配降下法

微分の連鎖計算 → backpropagation

NNの学習 (分類問題の場合)



勾配効果法 (GD) の実際

$$\Delta w_{ij}(t) = \alpha \Delta w_{ij}(t-1) - \epsilon \frac{\partial C}{\partial w_{ij}(t)}$$

momentum

学習係数
learning rate

サンプル数増 → 勾配の計算コスト増大

- 全サンプル使用 = batch
- 1サンプルのみ (オンライン学習) = 確率的 (stochastic GD)
- 中間 = "mini batch"

NN研究の歴史

NIPS2000では、タイトルにNNという語を含む論文は採択率と負の相関がある、とまで言われる [Simard03]

第1期

「冬の時代」

第2期

「冬の時代」

第3期

1960

1970

1980

1990

2000

2010

Perceptron
[Rosenblatt57]

Back-propagation
[Rumelhart+]

Convolutional NN
[LeCun+89]

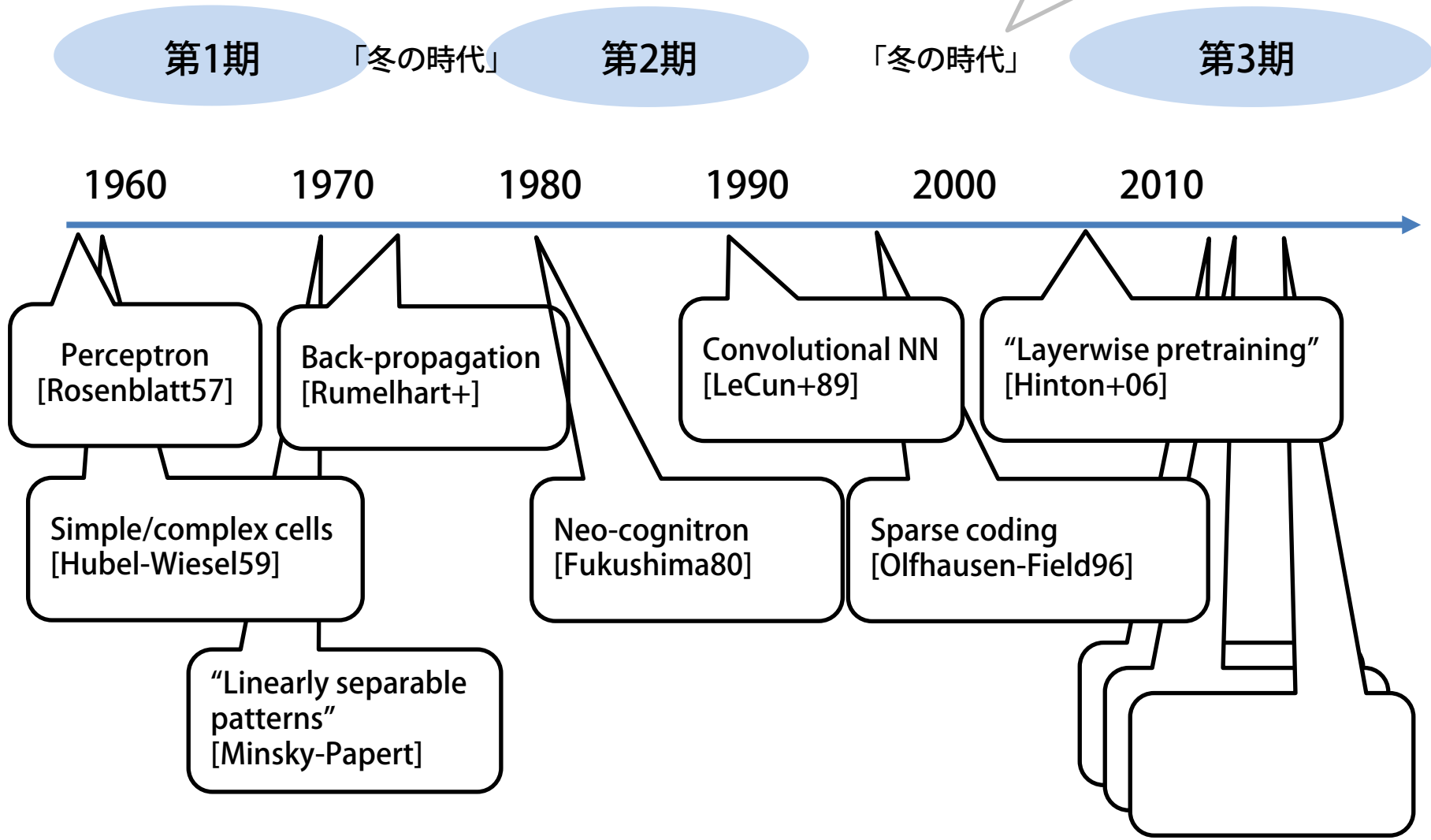
“Layerwise pretraining”
[Hinton+06]

Simple/complex cells
[Hubel-Wiesel59]

Neo-cognitron
[Fukushima80]

Sparse coding
[Olshausen-Field96]

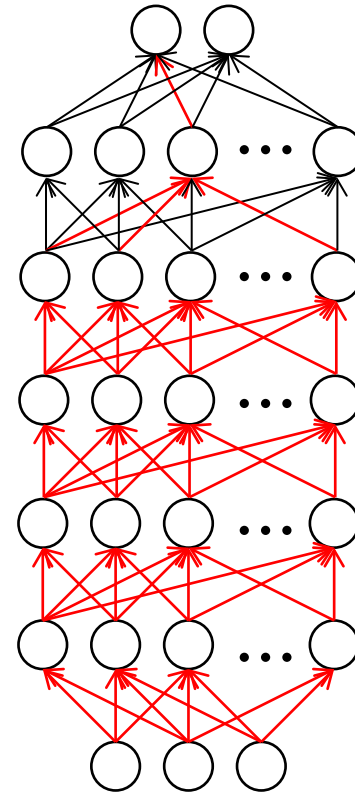
“Linearly separable
patterns”
[Minsky-Papert]



DNNの学習困難さ

Bengio, Learning Deep Architectures for AI, *Foundations and Trends in Machine Learning*, 2009

- **過学習**：学習誤差は小さくなるが
テスト誤差は小さくならない
 - 結線の重みをランダムに初期化し、
スタートした場合
- 経験的知見 = 「下層まで情報が伝わらない」
 - 「勾配が拡散する」
- 訓練データを表現するのに、上位層のみで事足りる



「冬の時代」の終焉

- DNNは学習が困難
 - 過学習

「層ごとのプレトレーニング」をやれば大丈夫！ [Hinton+06]

- 計算量
 - Backpropagation の計算量は膨大

計算能力向上；GPU や PC クラスターの登場



- 性能を引き出すのに必要なノウハウ（「黒魔術」）
 - 学習係数，モメンタム，ネットワーク構造など無数のパラメータ

???（進歩していない？）

NN研究の歴史

第1期

「冬の時代」

第2期

「冬の時代」

第3期

1960

1970

1980

1990

2000

2010

Perceptron
[Rosenblatt57]

Back-propagation
[Rumelhart+]

Convolutional NN
[LeCun+89]

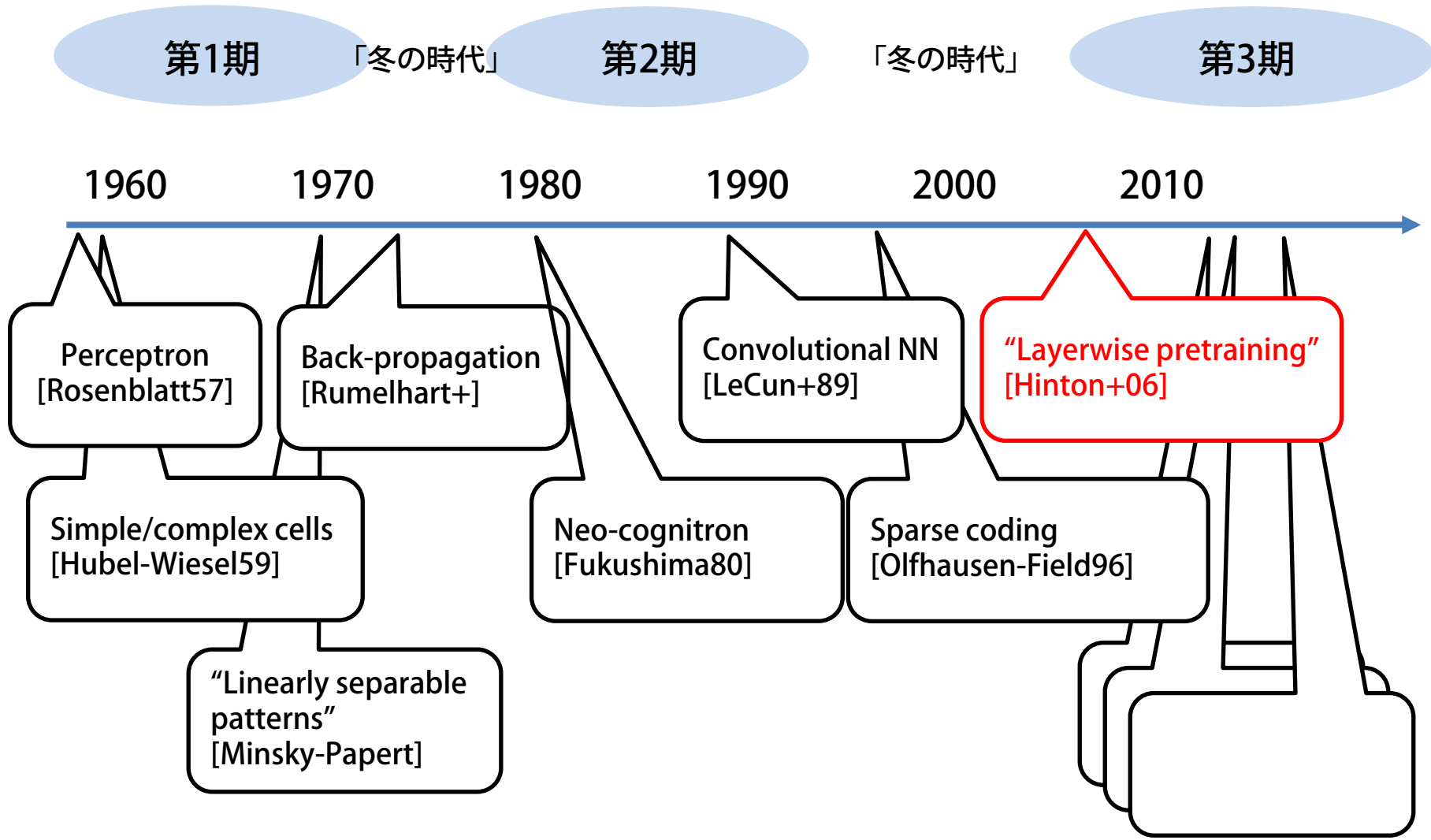
“Layerwise pretraining”
[Hinton+06]

Simple/complex cells
[Hubel-Wiesel59]

Neo-cognitron
[Fukushima80]

Sparse coding
[Olshausen-Field96]

“Linearly separable
patterns”
[Minsky-Papert]

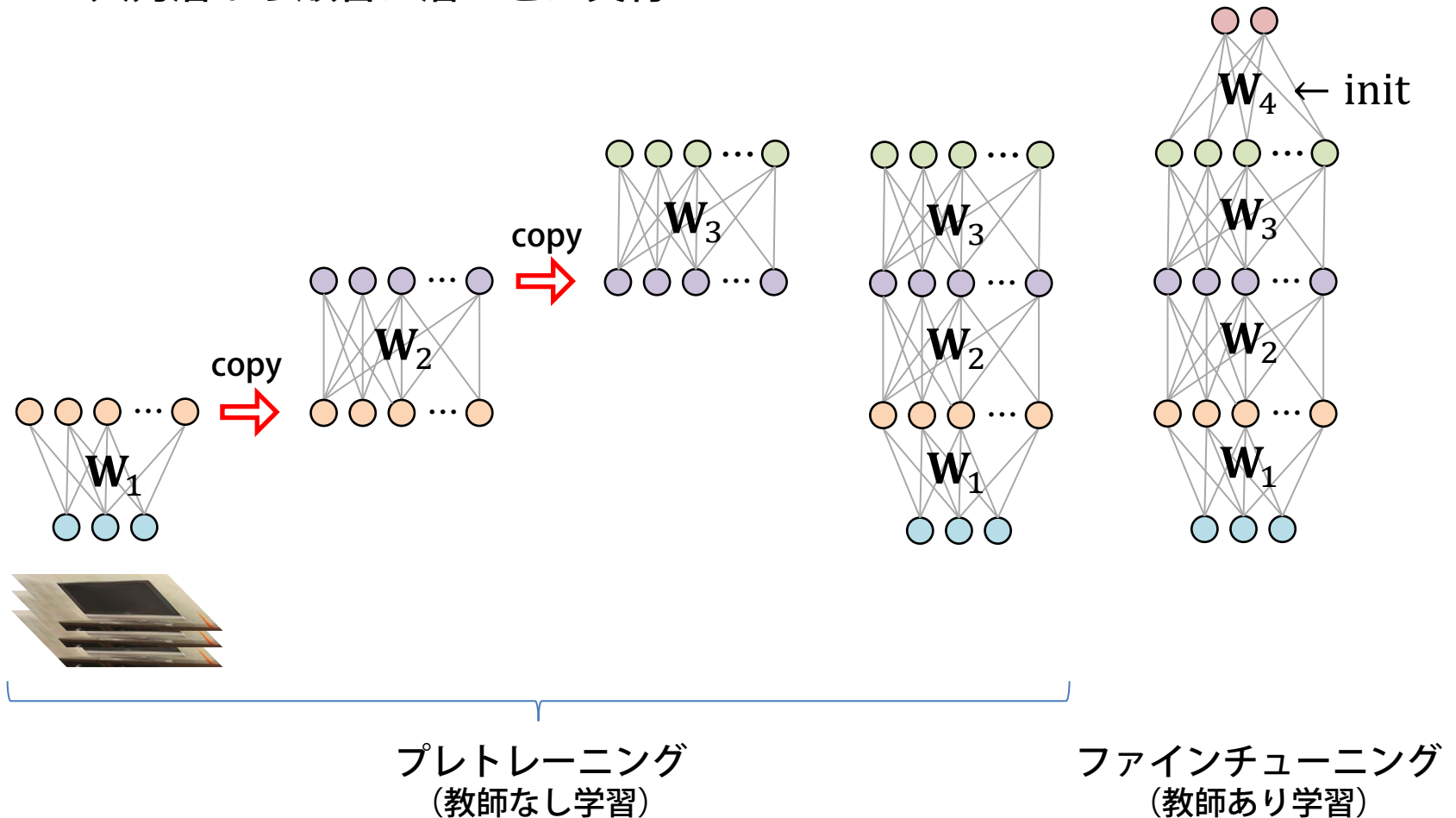


目次

- インTRODクシヨN
 - DNNの事例紹介
- NNの基礎と研究の略史
 - なぜNN研究は廃れたか？
- プレトレーニング
 - Encoder-decoder paradigm
- Convolutional Neural Networks (CNN)
 - プーリングと不変特徴
 - CNN vs. fully-connected NN
- ニューロサイエンス
 - どこまで分かっている？
- 生成モデル
 - Restricted Boltzmann Machine (RBM)
 - Deep Belief Network (DBN)
- 実践編
 - cuda-convnet を使ったMNISTの学習
- まとめ

DNNのプレトレーニング

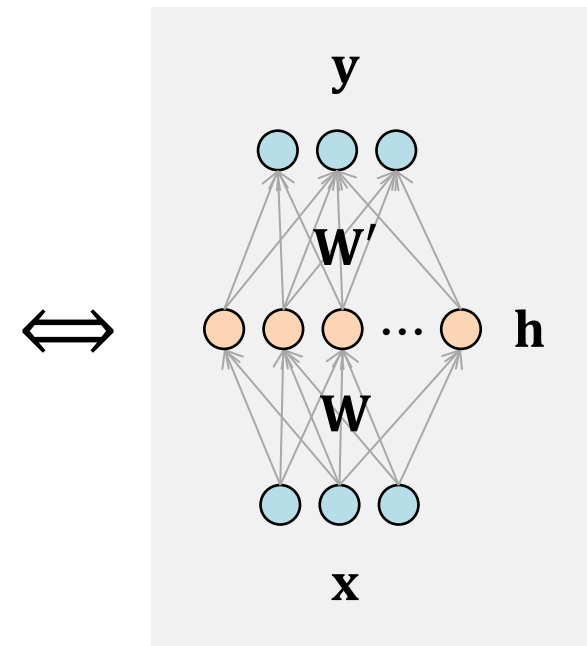
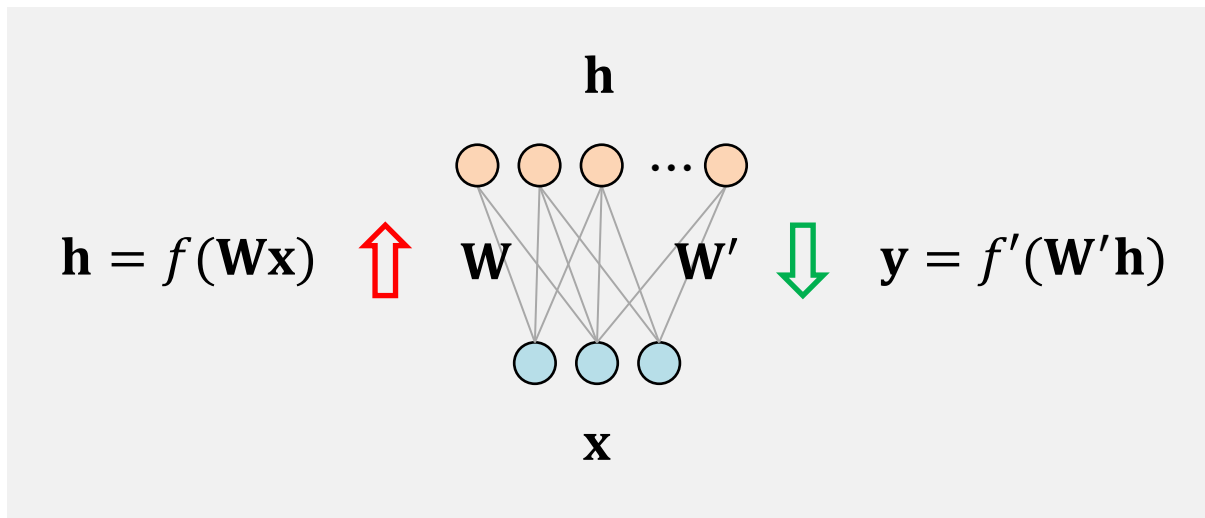
- プレトレーニング
 - 入力データの集合を再現できるように **教師無学習**
 - 入力層から順番に層ごとに実行



各層の教師無学習

- 入力データの集合を最もよく再現できるように学習
 - Encoder-decoder
- オートエンコーダ (Autoencoder)
 - もう一つの方法: Restricted Boltzmann Machine (RBM) → 「生成モデル」で説明

$$\min_{W, W'} \sum_i \| \mathbf{x}_i - \underbrace{f'}_{\text{green}}(\underbrace{W'}_{\text{red}} f(\mathbf{W} \mathbf{x}_i)) \|^2$$



プレトレーニングの効果

- プレトレーニングしておくとお学習が防げる
 - プレトレーニングで得たウェイトを初期値に
- 画像認識の state-of-the-artはすべて **Convolutional Neural Network**
 - プレトレーニング不要
 - 80年代にすでに多層の学習に成功していた！ [LeCun+89]

音声認識

[TABLE 1] COMPARISONS AMONG THE REPORTED SPEAKER-INDEPENDENT (SI) PHONETIC RECOGNITION ACCURACY RESULTS ON TIMIT CORE TEST SET WITH 192 SENTENCES.

METHOD	PER
CD-HMM [26]	27.3%
AUGMENTED CONDITIONAL RANDOM FIELDS [26]	26.6%
RANDOMLY INITIALIZED RECURRENT NEURAL NETS [27]	26.1%
BAYESIAN TRIPHONE GMM-HMM [28]	25.6%
MONOPHONE HTMS [29]	24.8%
HETEROGENEOUS CLASSIFIERS [30]	24.4%
MONOPHONE RANDOMLY INITIALIZED DNNs (SIX LAYERS) [13]	23.4%
MONOPHONE DBN-DNNs (SIX LAYERS) [13]	22.4%
MONOPHONE DBN-DNNs WITH MMI TRAINING [31]	22.1%
TRIPHONE GMM-HMMs DT W/ BMMI [32]	21.7%
MONOPHONE DBN-DNNs ON FBANK (EIGHT LAYERS) [13]	20.7%
MONOPHONE MCRBM-DBN-DNNs ON FBANK (FIVE LAYERS) [33]	20.5%
MONOPHONE CONVOLUTIONAL DNNs ON FBANK (THREE LAYERS) [34]	20.0%

画像認識

カテゴリ	データ	著者等	方法
一般物体認識	ILSVRC2012	Supervision	CNN w/o pretraining
	CIFAR10	Ciresan+12	CNN w/o pretraining
	NORB	Ciresan+12	CNN w/o pretraining
文字認識	「猫細胞」	Le+12	再構成TICA pretraining
	MNIST	Ciresan+12	CNN w/o pretraining
	HWDB1.0	Ciresan+12	CNN w/o pretraining

目次

- インTRODクシヨン
 - DNNの事例紹介
- NNの基礎と研究の略史
 - なぜNN研究は廃れたか？
- プレトレーニング
 - Encoder-decoder paradigm
- Convolutional Neural Networks (CNN)
 - プーリングと不変特徴
 - CNN vs. fully-connected NN
- ニューロサイエンス
 - どこまで分かっている？
- 生成モデル
 - Restricted Boltzmann Machine (RBM)
 - Deep Belief Network (DBN)
- 実践編
 - cuda-convnet を使ったMNISTの学習
- まとめ

Convolutional Neural Network (CNN)

第1期

「冬の時代」

第2期

「冬の時代」

第3期

1960

1970

1980

1990

2000

2010

Perceptron
[Rosenblatt57]

Back-propagation
[Rumelhart+]

Convolutional NN
[LeCun+89]

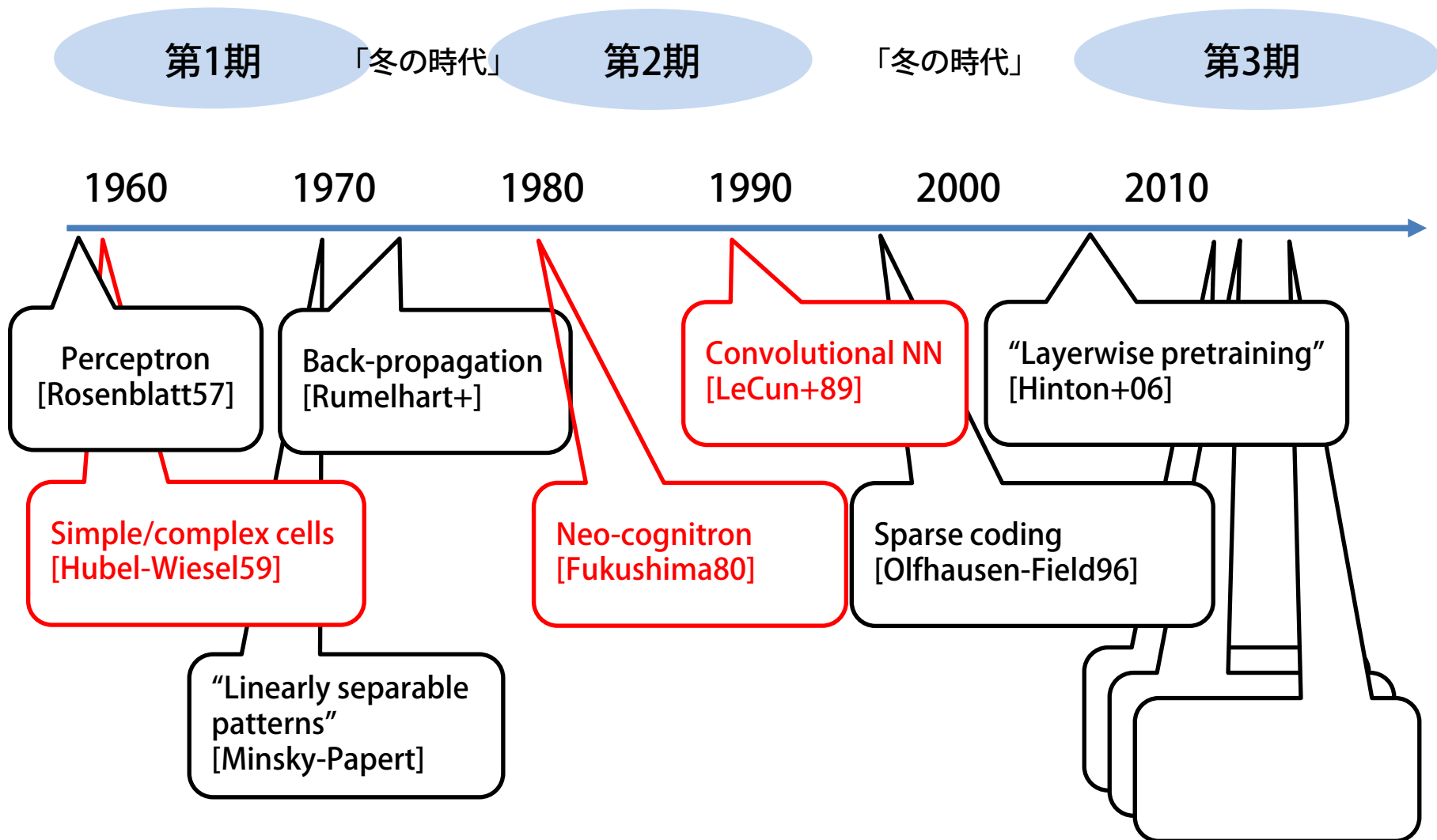
“Layerwise pretraining”
[Hinton+06]

Simple/complex cells
[Hubel-Wiesel59]

Neo-cognitron
[Fukushima80]

Sparse coding
[Olshausen-Field96]

“Linearly separable
patterns”
[Minsky-Papert]



Convolutional Neural Network (CNN)

- Neocognitronにルーツ [Fukushima80]
- Backpropagationによる教師有学習と手書き文字認識への応用 [LeCun+89]
 - Backpropagation Applied to Handwritten Zip Code Recognition, *Neural Computation*, 1989
- 神経科学の知見が基礎
 - Hubel-Wiesel の単純細胞・複雑細胞
 - 局所受容野 (local receptive field)

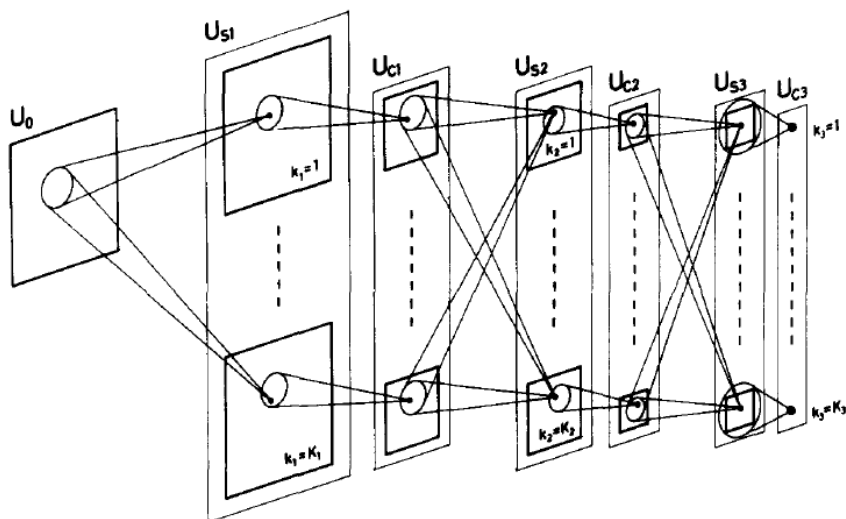


Fig 4 Schematic diagram illustrating the interconnections between layers in the neocognitron

[Fukushima+83]

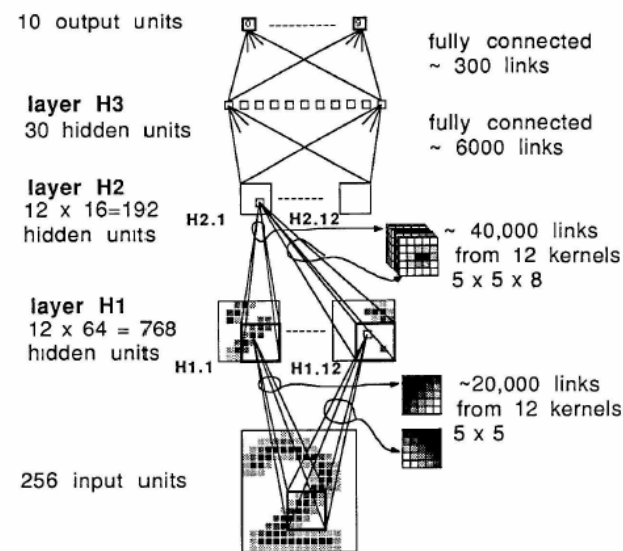
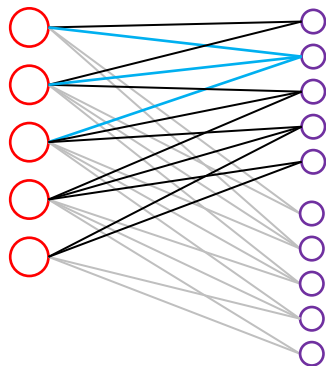
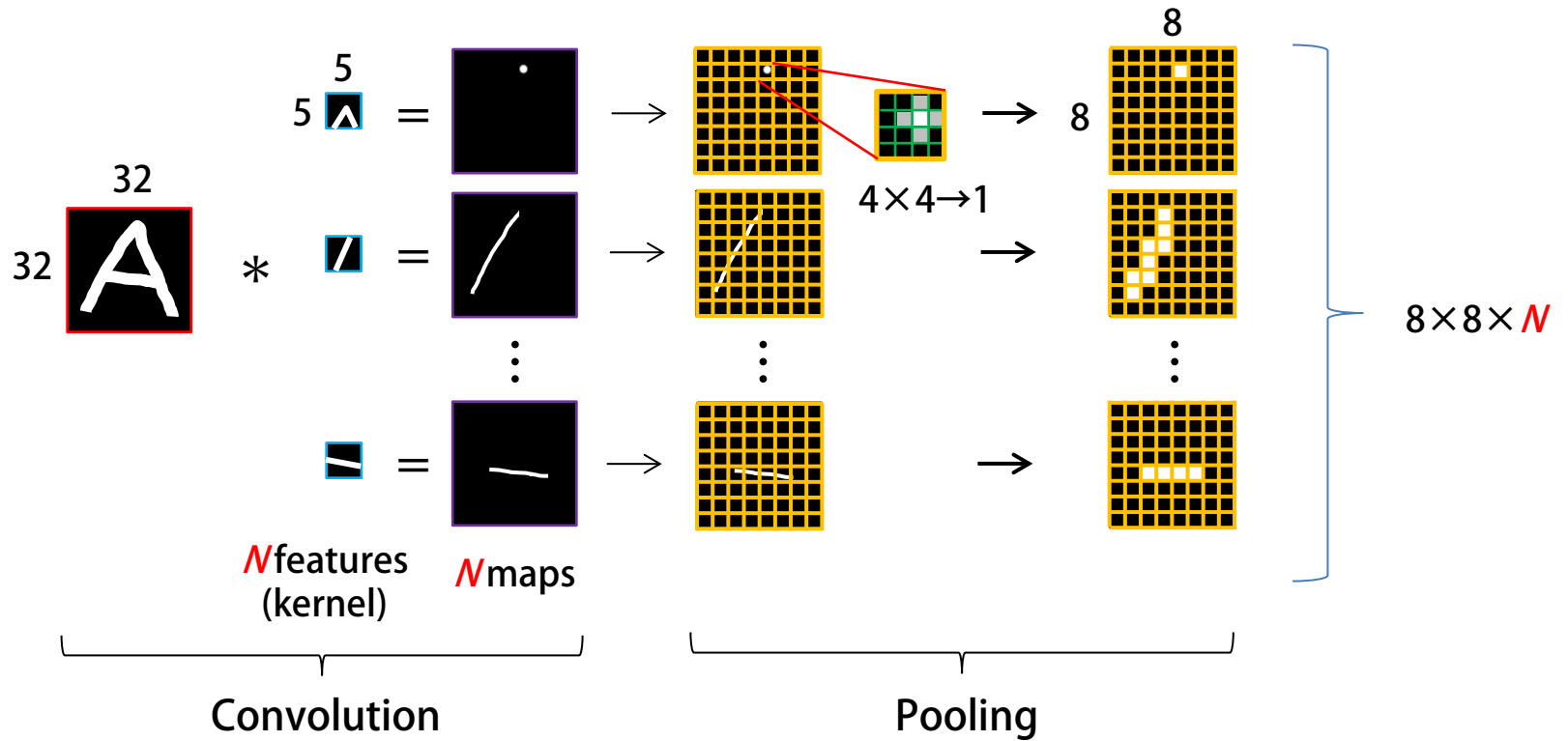


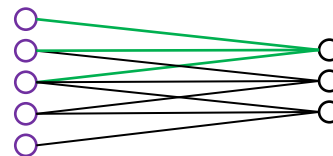
Figure 3 Log mean squared error (MSE) (top) and raw error rate (bottom) versus number of training passes

[LeCun+89]

Convolutional Neural Network (CNN)

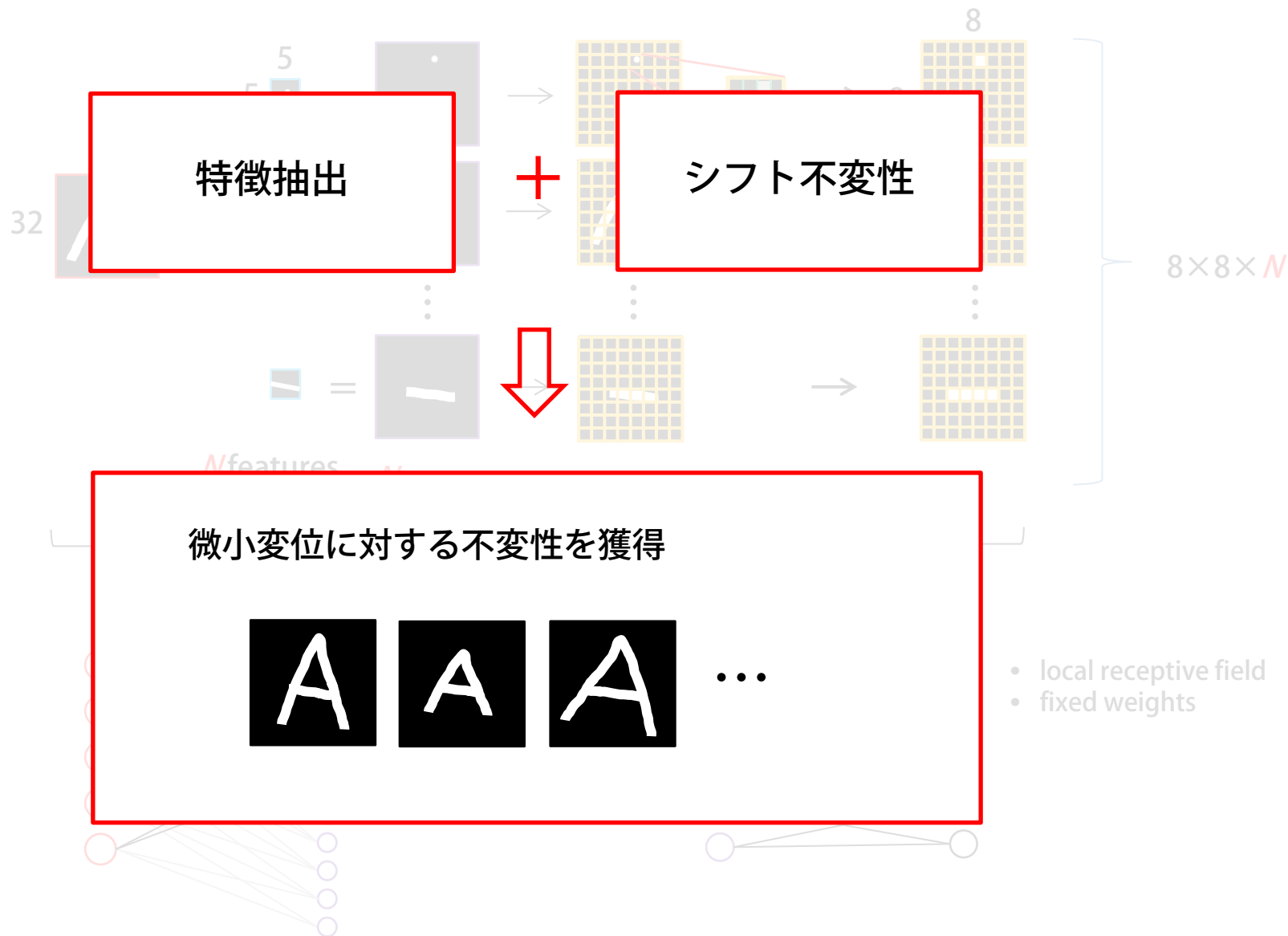


- local receptive field
- tied weights



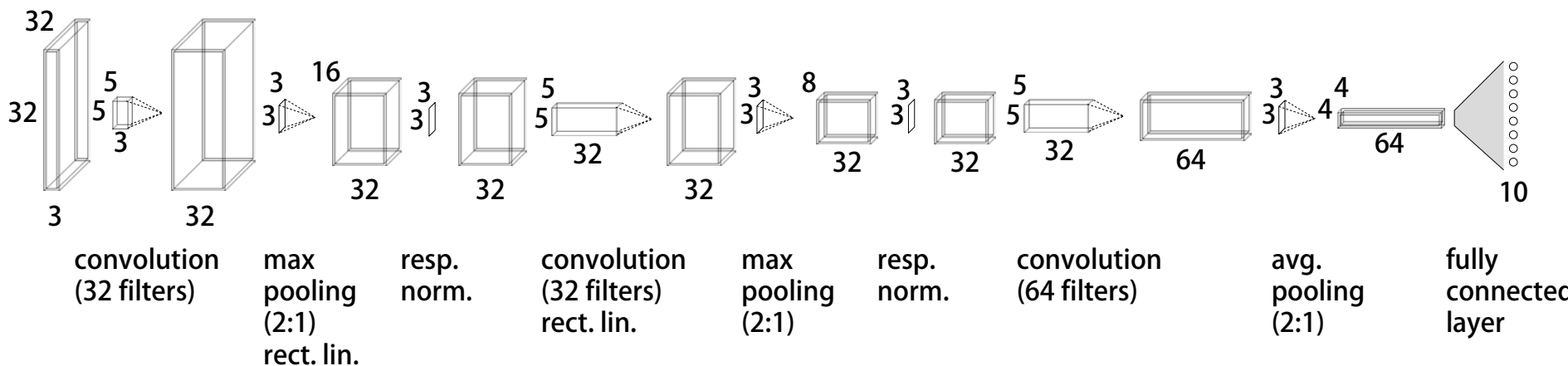
- local receptive field
- fixed weights

Convolutional Neural Network (CNN)



Convolutional Neural Network (CNN)

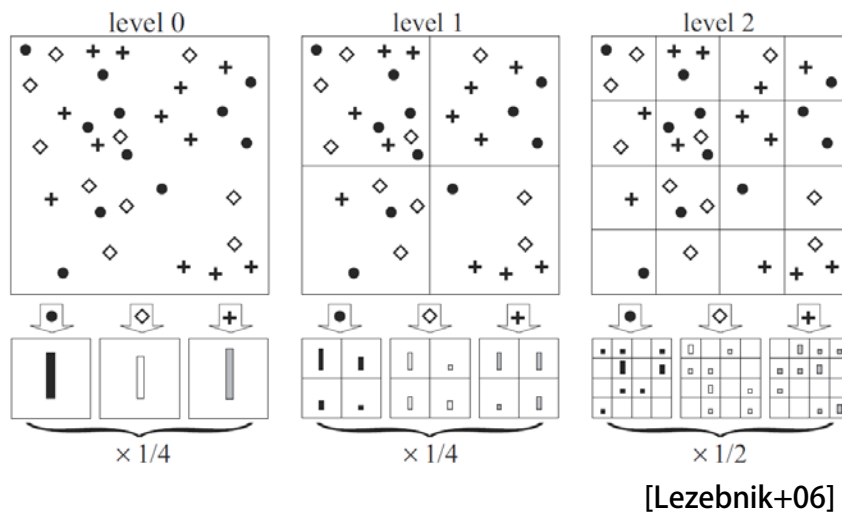
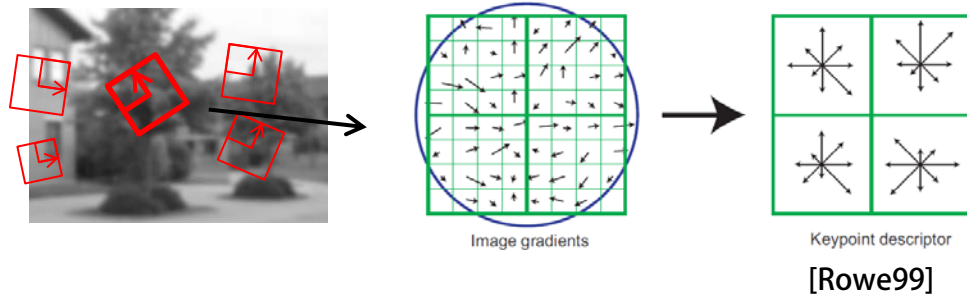
- Convolution + pooling を繰り返す (= Deep CNN) ことで, 多様な変形に対する不変性を獲得
- 学習可能
 - 対象: フィルタの係数, 方法: バックプロパゲーション



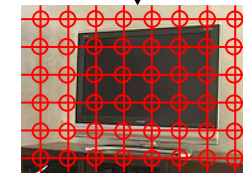
CIFAR10用 CNN [Krizhevsk+11]

非NNの方法とプーリング

- 一般物体認識でおなじみ
 - SIFT [Rowe99] (ディスクリプタの生成)
 - Bag of Features
 - Spatial pyramid [Lezebnik+06]



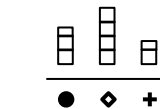
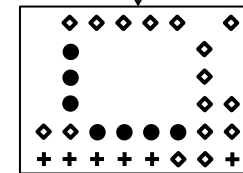
Image



Local features



Visual words
(Vocabulary)



Global feature

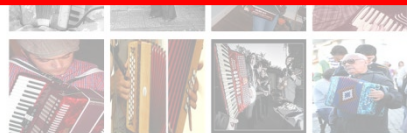
事例 2: 一般物体認識 (Hintonのグループ)

Krizhevsk et al., ImageNet Classification with Deep Convolutional Neural Networks, NIPS2012

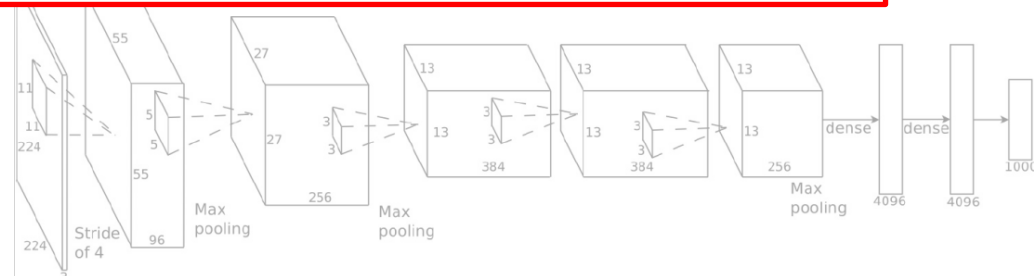
- IMAGENET Large Scale Visual Recognition Challenge 2012
 - 1000カテゴリ・カテゴリあたり約1000枚の訓練画像

「さらなるデータと、もっと早いGPUが手に入れば、ただちに一層の性能向上が可能」 [Krizhevsky+12]

- ネットワーク大規模化 → 潜在的な認識能力向上 + 過学習リスク増大
- 大量データ → 過学習のリスクを低減できる + 計算時間がかかる



“accordion ...”



7-layer NN

CNN vs. Fully-connected NN

CNN

- プレトレーニング不要
 - local RF, tied weights
 - “prewired”
- アーキテクチャ設計の難しさ
 - フィルタサイズ・ストライド・マップ数
 - プーリングサイズ・ストライド

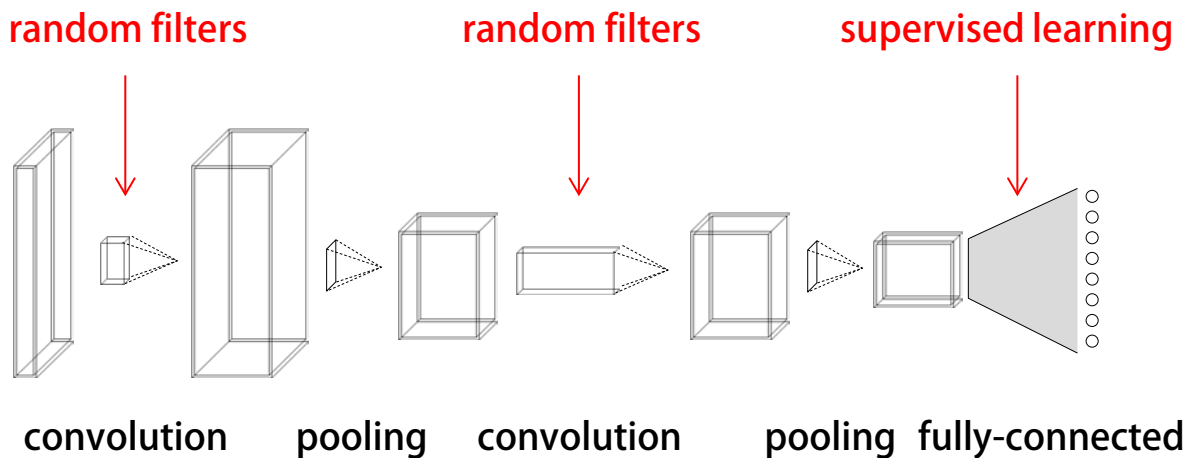
Fully-connected NN

- プレトレーニング有効
 - 以前は必須と考えられていたが、そうでもないらしい
- 別の方法
 - Drop-out: Fully-connect NN の過学習を避ける学習方法 [Hinton12]
 - discriminative プレトレーニング

ランダムフィルタ：アーキテクチャの重要性

Jarrett et al., What is the best multi-stage architecture for object recognition? ICCV09

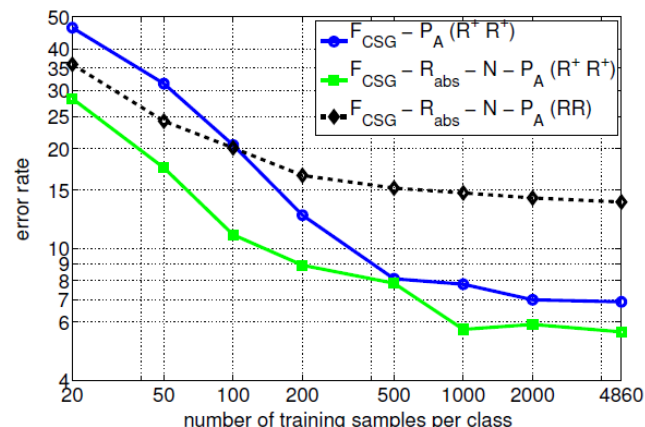
- フィルタをランダムとし，最上位層の fully-connected 層のみ学習



Caltech-101

アーキテクチャ	ランダム フィルタ	フィルタも 学習
2層，絶対値プーリング	62.9%	64.7%
2層，平均プーリング	19.6%	31.0%
1層，絶対値プーリング	53.3%	54.8%

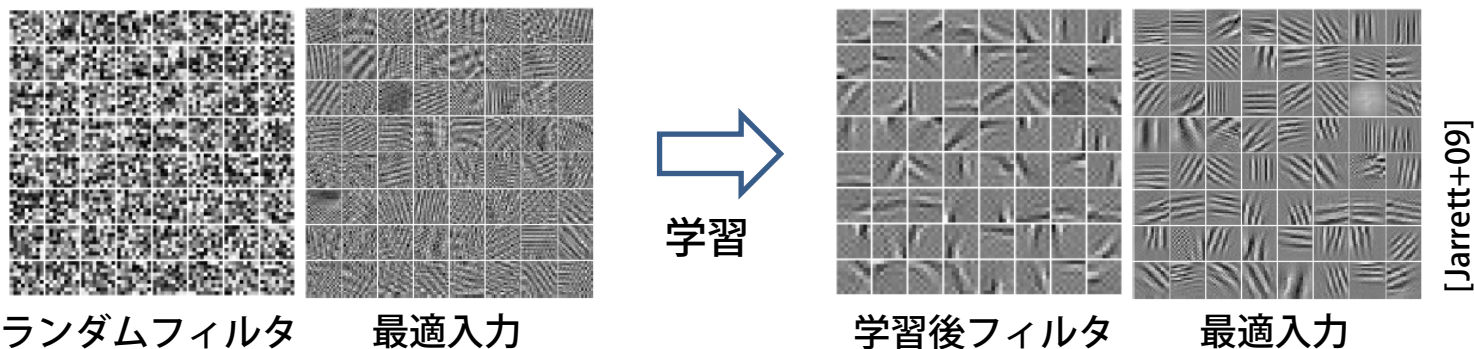
NORB



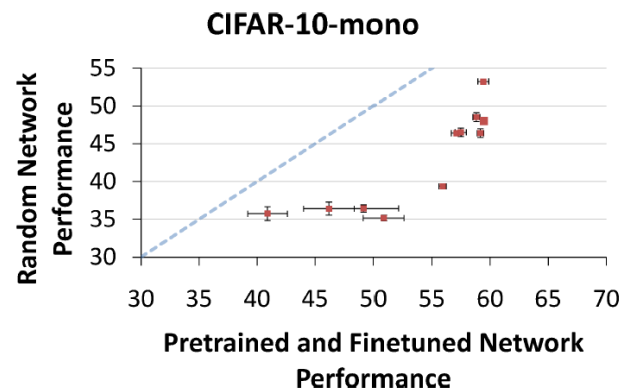
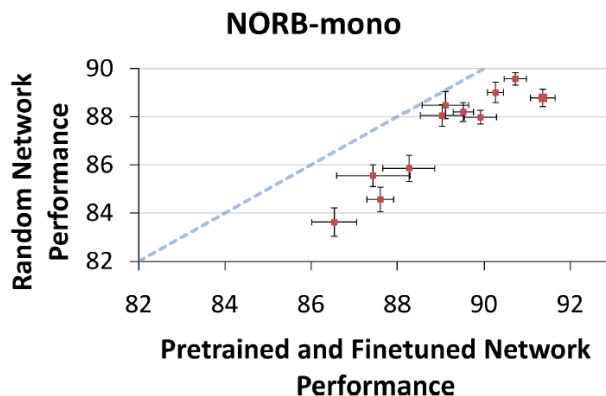
ランダムフィルタ：アーキテクチャの重要性

Saxe et al., On random weights and unsupervised feature learning, ICML2010

- 学習アルゴリズムよりもアーキテクチャがずっと大事
- プーリング層のユニットが最も反応する最適入力：
 - 理論的説明 [Saxe+10]



- アーキテクチャの性能予測をランダムフィルタで [Saxe+10]
 - アーキテクチャ探索時間を節約可

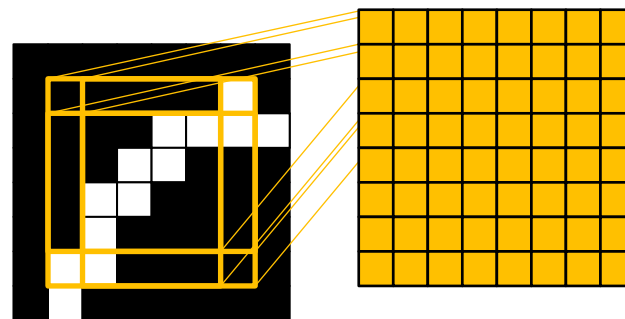
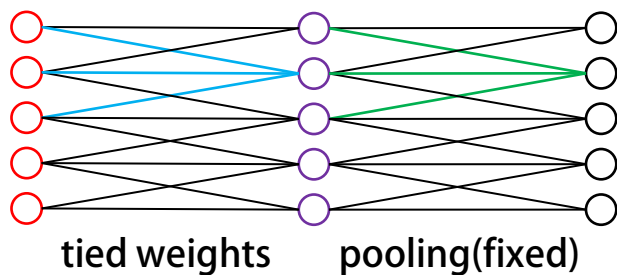


CNNの改良例：Tiled CNN

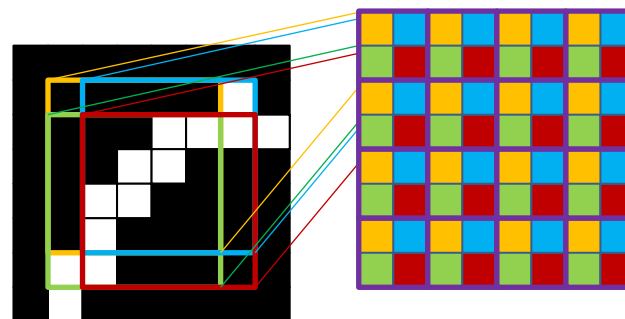
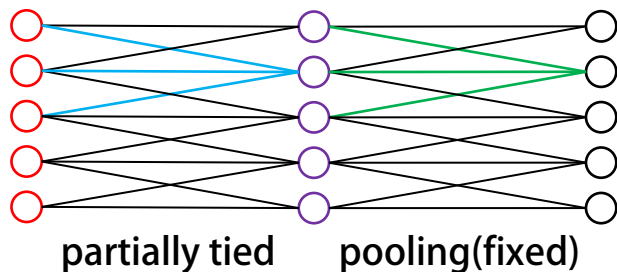
Le et al., Tiled convolutional neural networks, NIPS2010

- CNNでは単純なシフト不変性しか獲得できない
 - 局所受容野 + tied weights
- Tiled CNN
 - 場所ごとに違うフィルタ ← より複雑な変形に対する不変性の獲得
 - Topographic ICA (TICA)

CNN



Tiled CNN



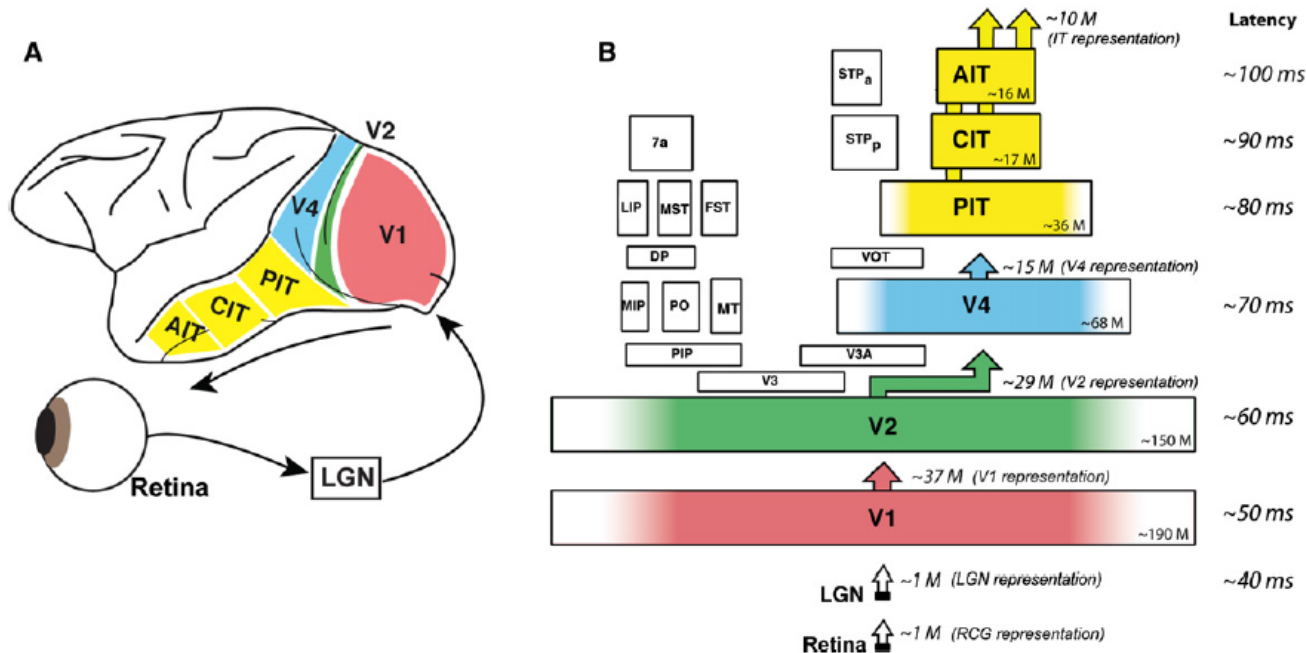
目次

- インTRODクシヨN
 - DNNの事例紹介
- NNの基礎と研究の略史
 - なぜNN研究は廃れたか？
- プレトレーニング
 - Encoder-decoder paradigm
- Convolutional Neural Networks (CNN)
 - プーリングと不変特徴
 - CNN vs. fully-connected NN
- ニューロサイエンス
 - どこまで分かっている？
- 生成モデル
 - Restricted Boltzmann Machine (RBM)
 - Deep Belief Network (DBN)
- 実践編
 - cuda-convnet を使ったMNISTの学習
- まとめ

脳の視覚情報処理

DiCarlo, Zoccolan, Rust, How does the brain solve visual object recognition?, Neuron, 2012

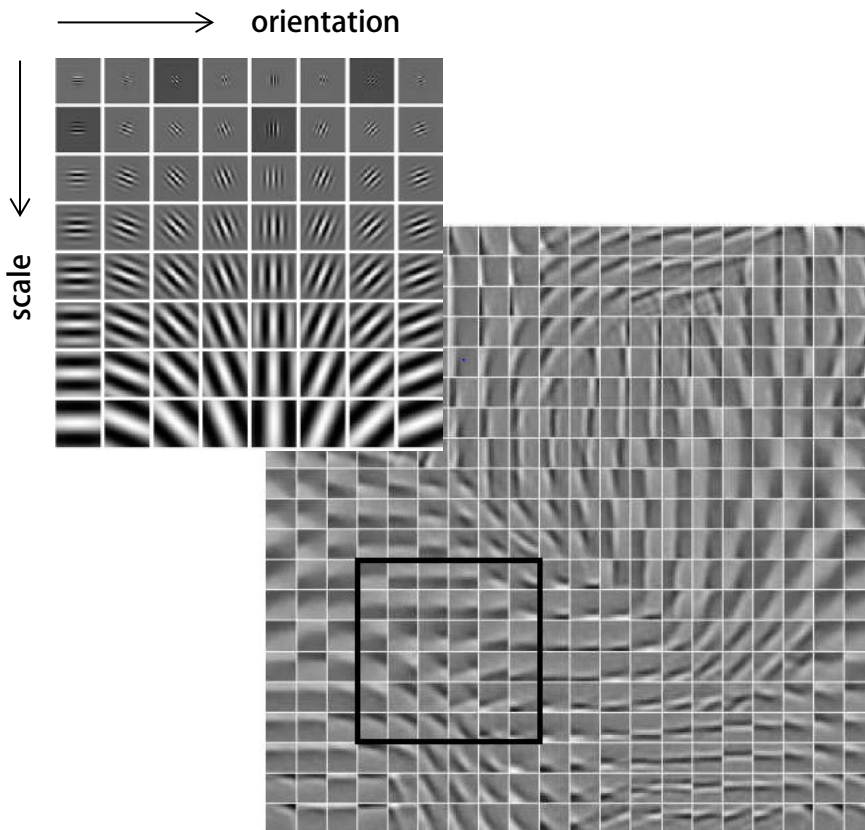
- 視覚野（腹側皮質視覚路）の構造
 - フィードフォワードで伝播
 - 階層性：単純な特徴抽出 → 複雑なものへ
 - IT野での処理：「おばあさん細胞」仮説と「population code」仮説



[Dicarlo+12]

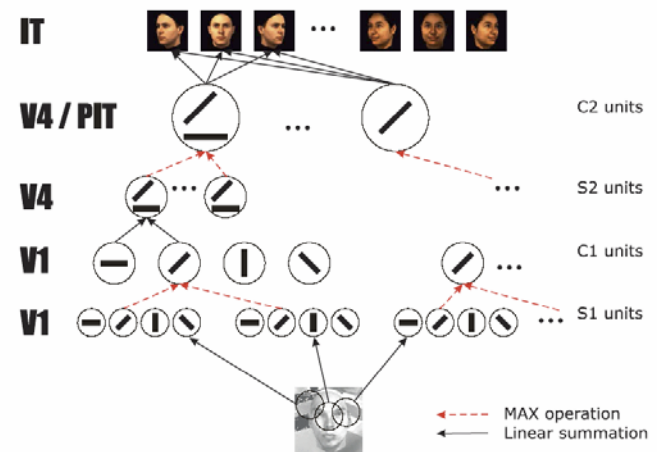
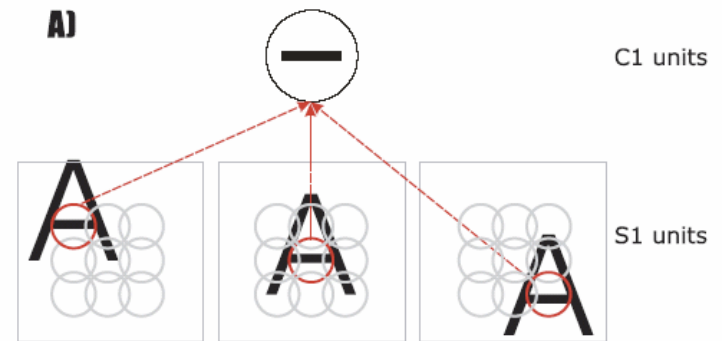
V1

- ガボールウェーブレット
 - 位置 / 向き / スケール
 - Topographic map



Kavukcuoglu, Ranzato, Fergus, LeCun, Learning Invariant Features through Topographic Filter Maps, CVPR09

- Simple cells/complex cells [Huber-Wiesel59]



Serre et al, Object Recognition with Features Inspired by Visual Cortex, CVPR05

スパースコーディング

Olshausen, Field, Sparse Coding with an Overcomplete Basis Set: a Strategy Employed by V1?, Vis. Research96

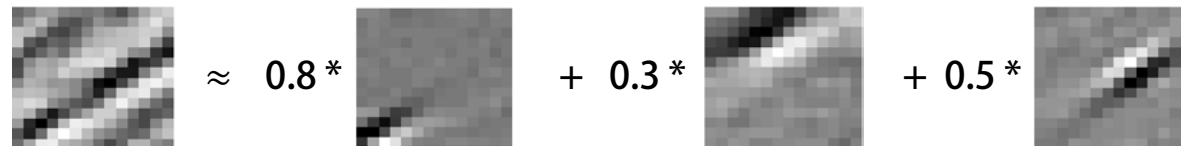
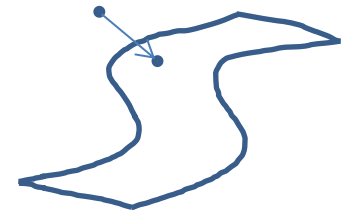
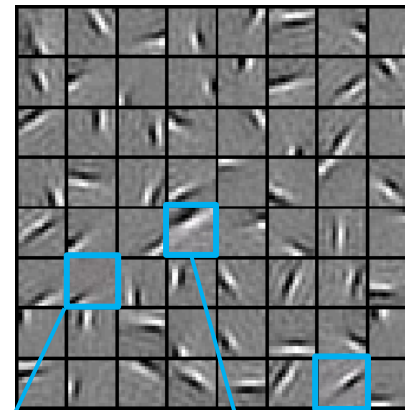
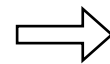
$$\min_{\mathbf{D}, \mathbf{h}} \sum_i \frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\mathbf{h}_i\|_2^2 + \lambda \|\mathbf{h}_i\|_1$$

過完備性
(overcomplete)

Natural image patches $\{\mathbf{x}_i\}$



$\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_n]^T$



スパースコード $\mathbf{h} = [0, 0, \dots, 0, 0.8, 0, \dots, 0, 0.3, 0, \dots, 0, 0.5, \dots]$

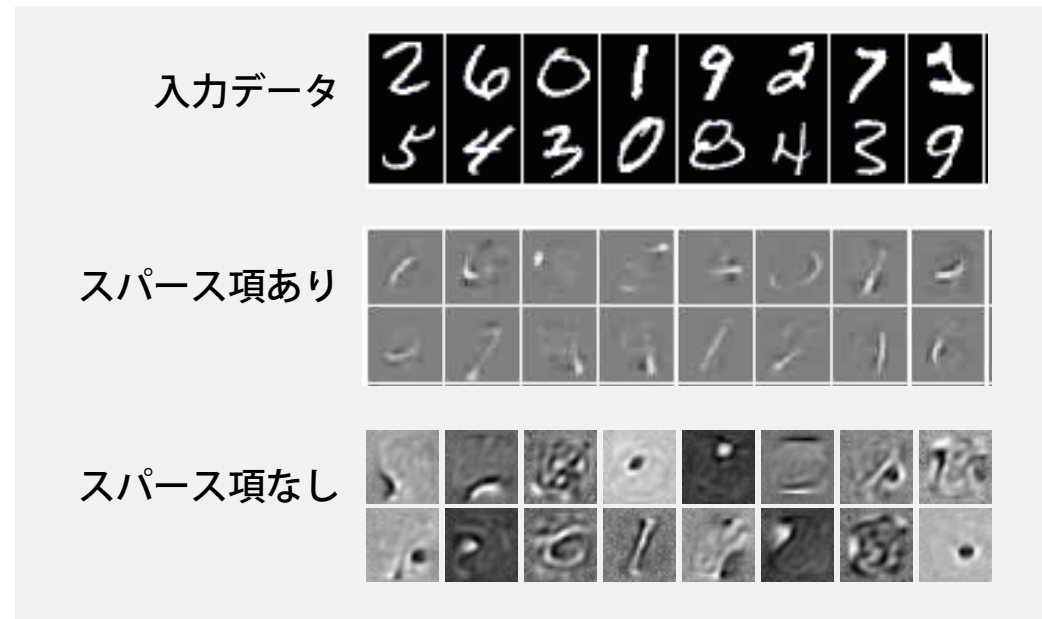
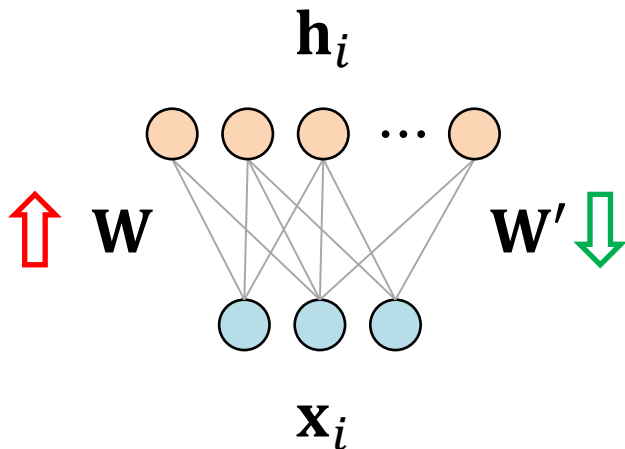
スパースオートエンコーダ

Ranzato et al., Efficient learning of sparse representations with an energy-based model, NIPS06 他

- 出力層のユニットの反応がスパースになるように正則化
 - 過完備性

$$\min_{W, W'} \sum_i \| \mathbf{x}_i - \underbrace{f'}(\underbrace{W' f(W \mathbf{x}_i)}) \|^2 + \sum_i \underbrace{g(f(W \mathbf{x}_i))}$$

e.g. $\frac{1}{n} \sum_i \mathbf{h}_i \sim \begin{bmatrix} \epsilon \\ \vdots \\ \epsilon \end{bmatrix}$

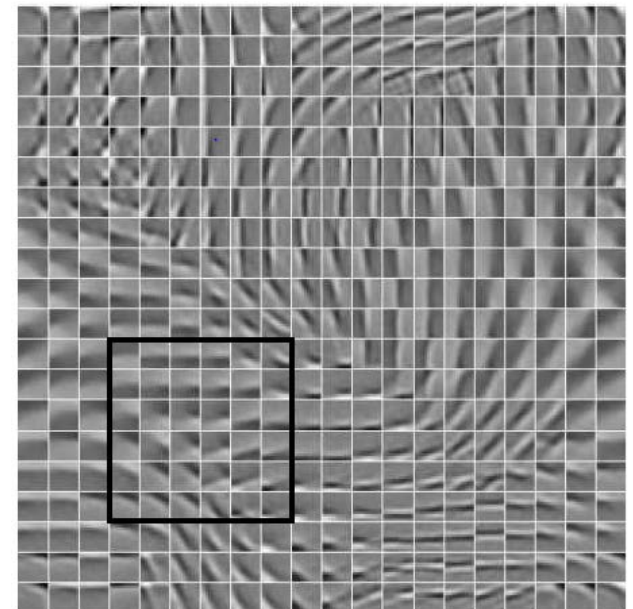
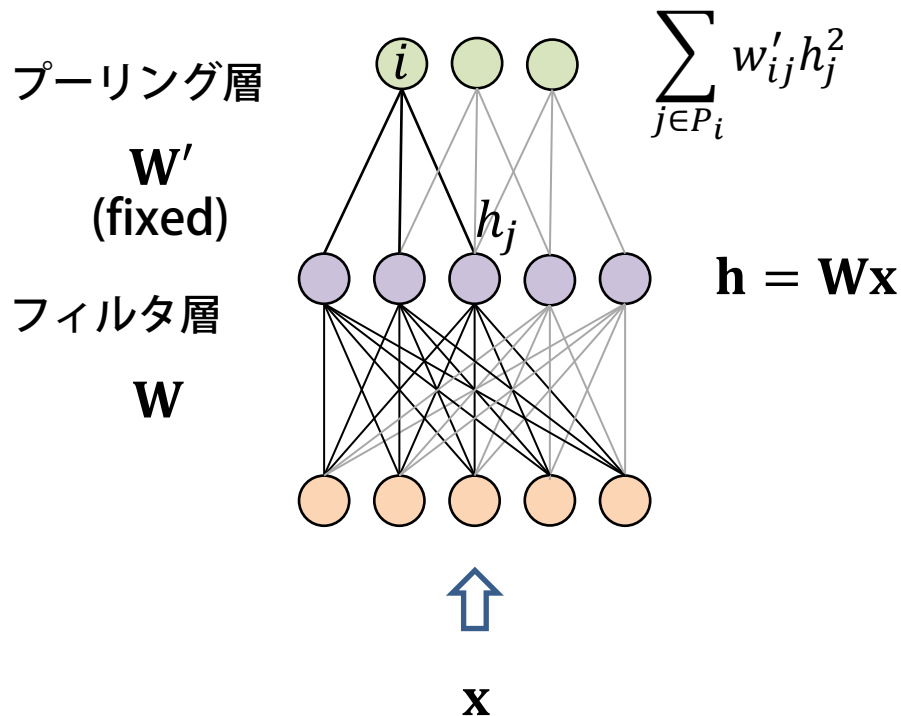


Topographic ICA

Hyvarinen and Hoyer, A two-layer sparse coding model learns simple and complex cell..., Vision Research, 2001

- 隣接ユニットの特徴（フィルタ）が類似するように

$$\min_{\mathbf{W}} \sum_i \sqrt{\sum_{j \in P_i} w'_{ij} h_j^2}, \quad s.t. \mathbf{W}^T \mathbf{W} = \mathbf{I}$$

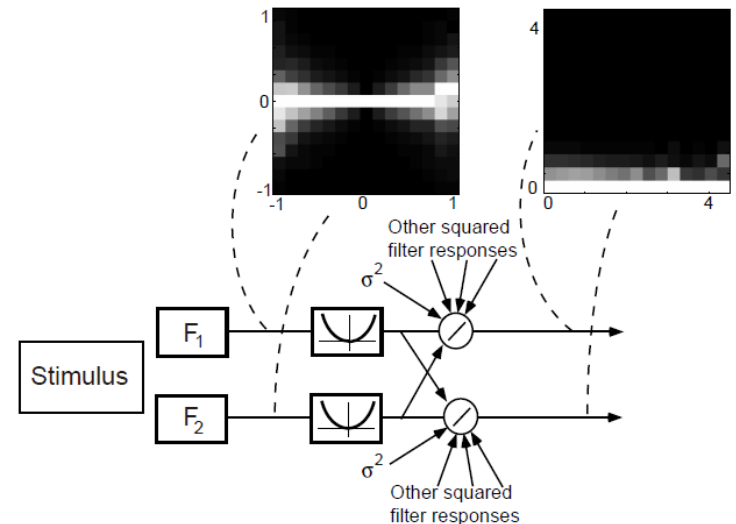
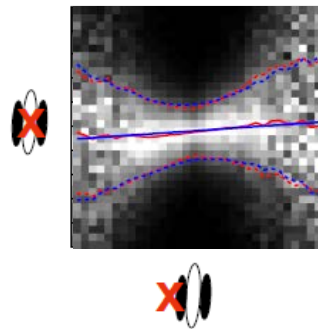
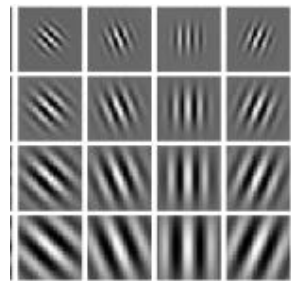
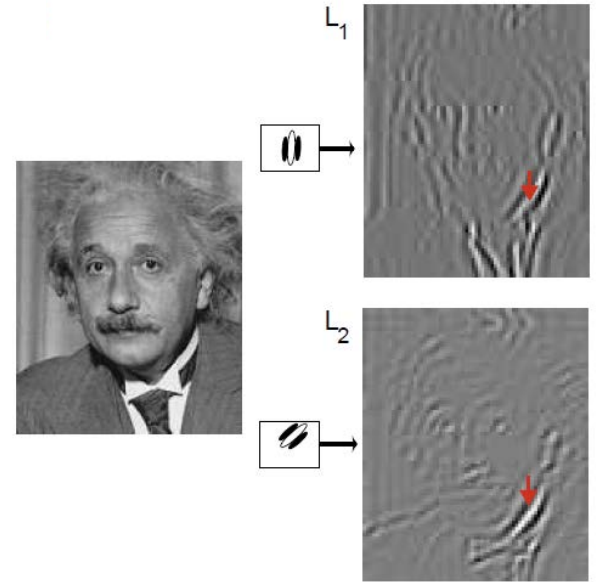
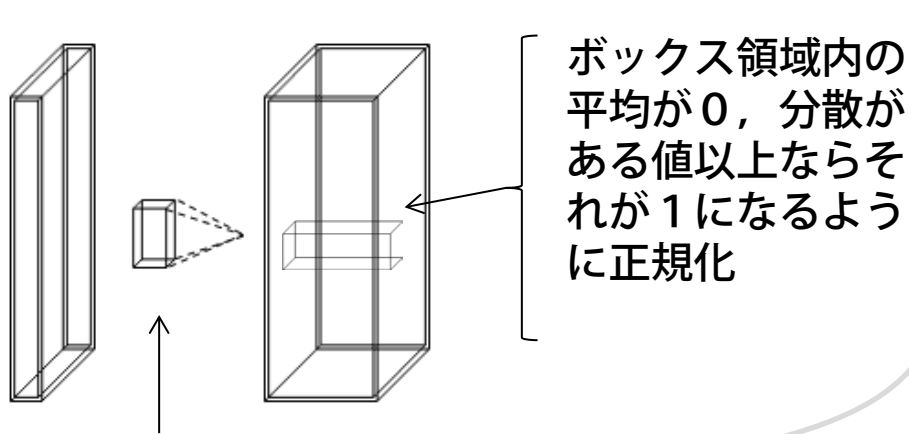


[Kavukcuoglu09]

局所コントラスト正規化

Schwartz, Simoncelli, Natural signal statistics and sensory gain control, Nature neuroscience, 2001

- 視覚野ニューロンの挙動のモデル
 - 画像認識でも効果
 - 自然画像統計に基づく説明 [Schwartz-Simoncelli01]

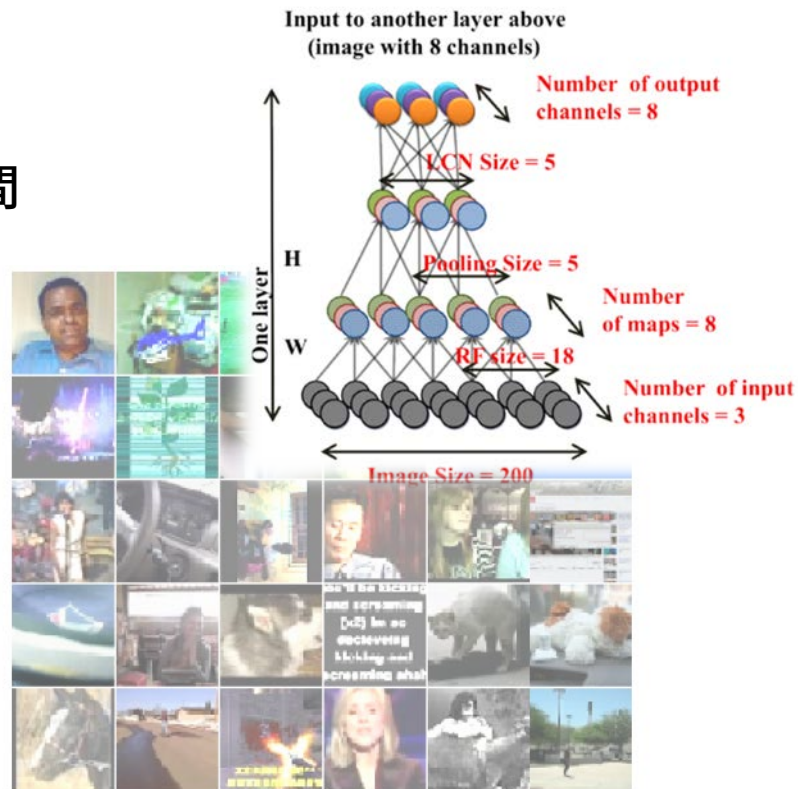


[Schwartz-Simoncelli01]

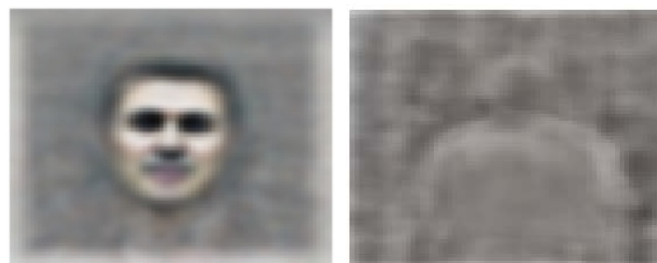
事例 1 : 画像特徴の無教師学習 (Ng のグループ)

Le et al., Building High-level Features Using Large Scale Unsupervised Learning, ICML2012

- 12層NNを使った無教師学習
 - パラメータ数10億個!
 - 16コアPC1000台のPCクラスター×3日間
 - YouTubeの画像1000万枚
- 「おばあさん細胞」の生成を確認



The New York Times (2012/6/25)



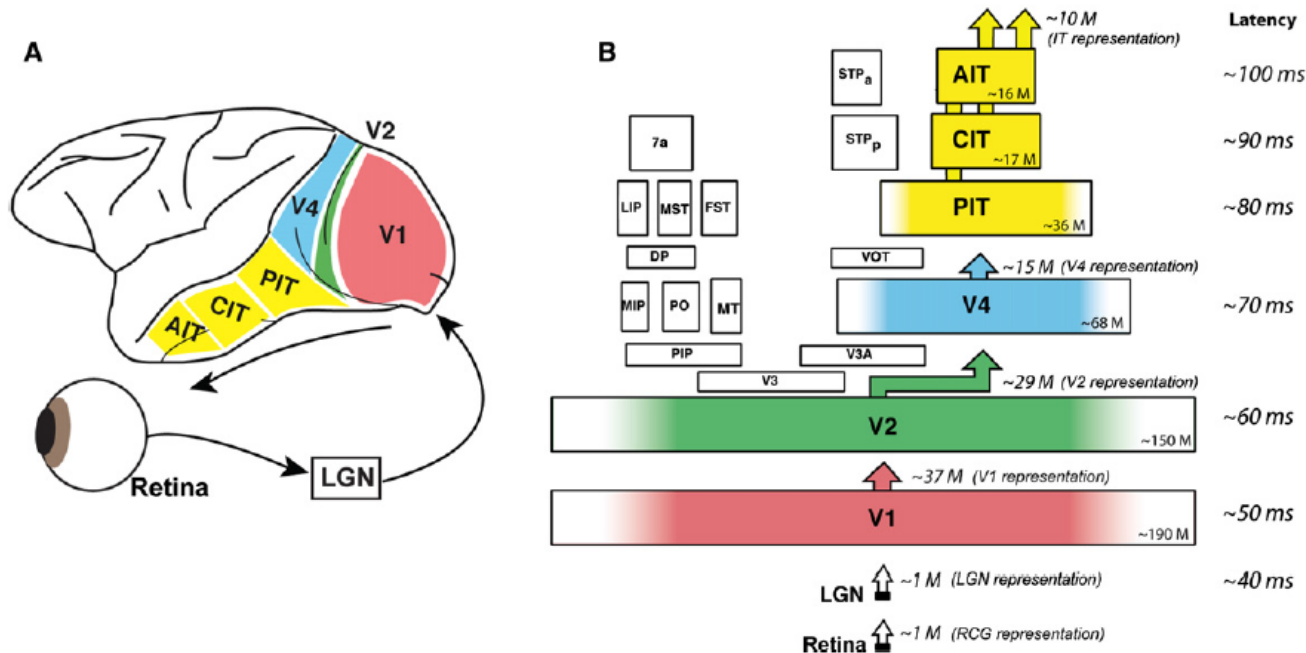
顔

人の体

脳の視覚情報処理

DiCarlo, Zoccolan, Rust, How does the brain solve visual object recognition?, Neuron, 2012

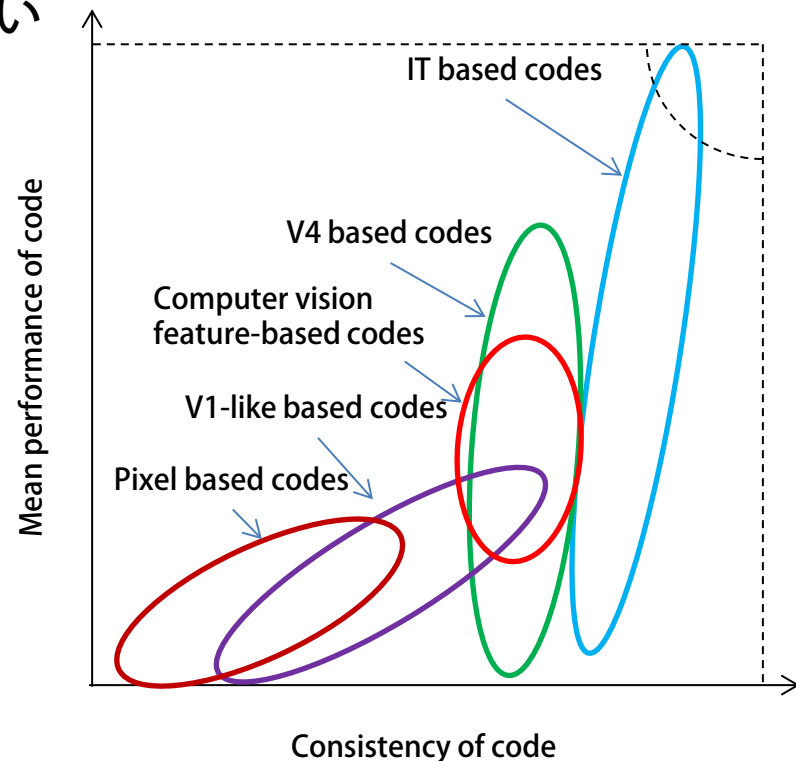
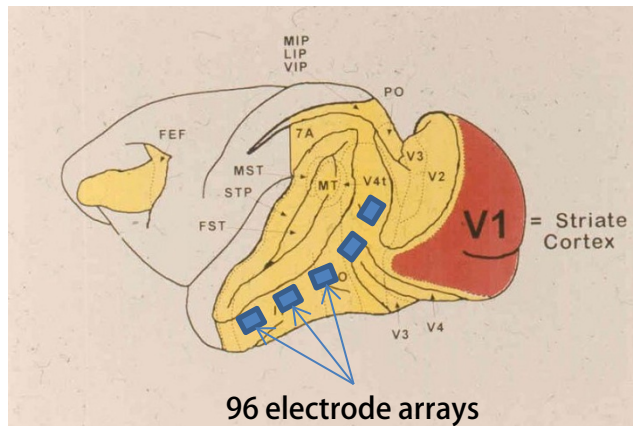
- 視覚野（腹側皮質視覚路）の構造
 - フィードフォワードで伝播
 - 階層性：単純な特徴抽出 → 複雑なものへ
 - IT野での処理：「おばあさん細胞」仮説と「population code」仮説



[Dicarlo+12]

一般物体認識は脳でどのように行われるか？

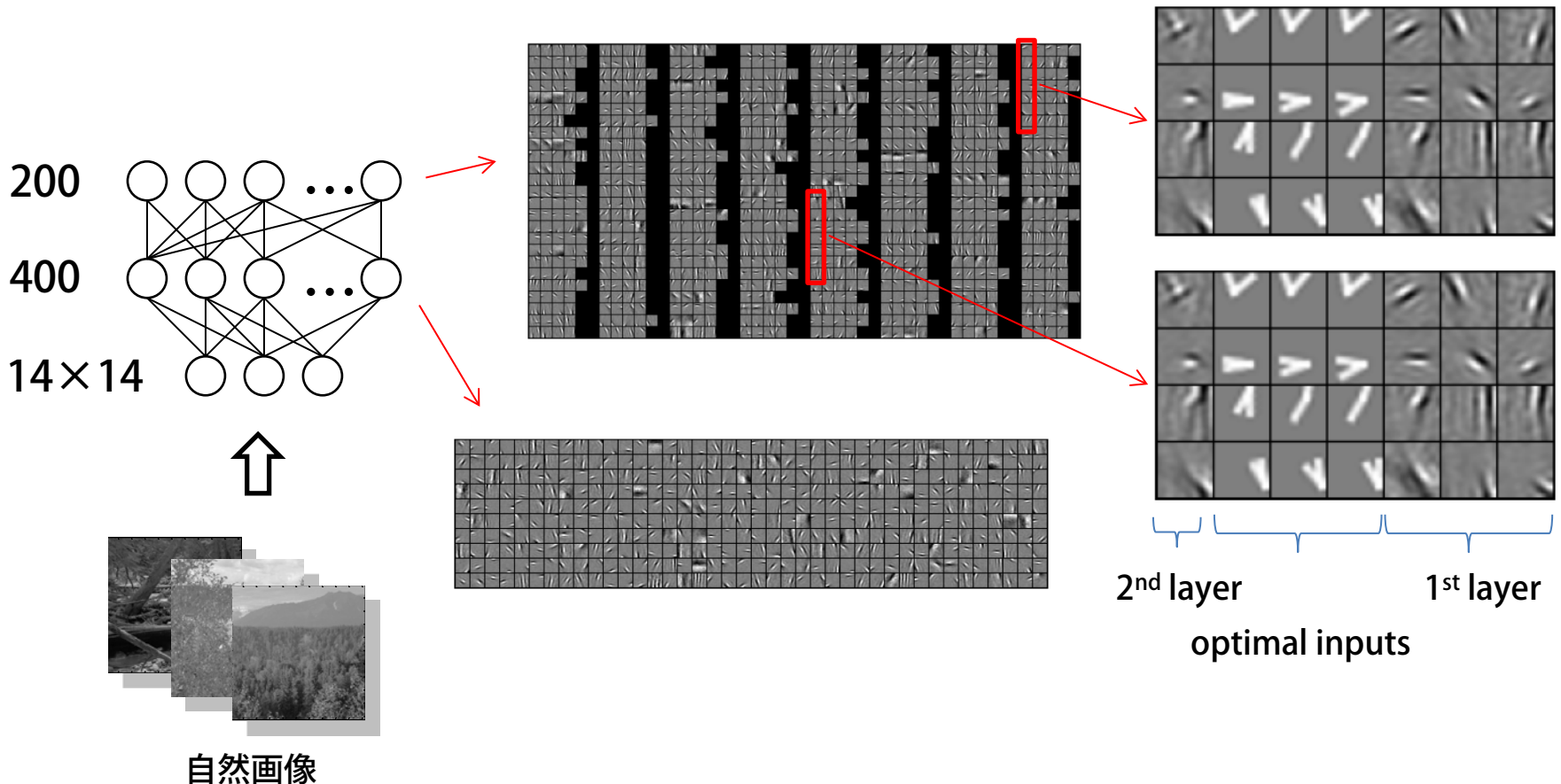
- “The brain’s bag of features is better than ours.”
- サルに画像を提示；脳のV1,V2,V4,IT野の活動パターンを計測
- IT野の計測パターン由来の“code”をSVMで分類すると，人と同じ認識性能を達成できる
 - V4野までの“code”ではできない



V2 のモデル

Lee et al., Sparse deep belief net model for visual area V2, NIPS08

- スパースRBM を使った V2 のモデル
 - 現実のV2ニューロンの応答[Ito-Komatsu04]を再現
 - V1より複雑な形（アングルやジャンクション）に反応



目次

- インTRODクシヨン
 - 事例紹介
- NNの基礎と研究の略史
 - なぜNN研究は廃れたか？
- DNNのプレトレニング
 - オートエンコーダ
- Convolutional Neural Networks (CNN)
 - プーリングと不変特徴
 - ネットワークアーキテクチャの重要性
 - CNNの改良
 - プレトレニングの必要性
- ニューロ・ビジョンサイエンスのアプローチ
 - 明らかになっていること
- 生成モデル
 - Restricted Boltzmann Machine (RBM), Deep Belief Network (DBN)
- 実践編 (ソフトウェア紹介)
 - Theano
 - cuda-convnet
- まとめ

生成モデルを用いたディープラーニング

齋藤 真樹(東北大学)

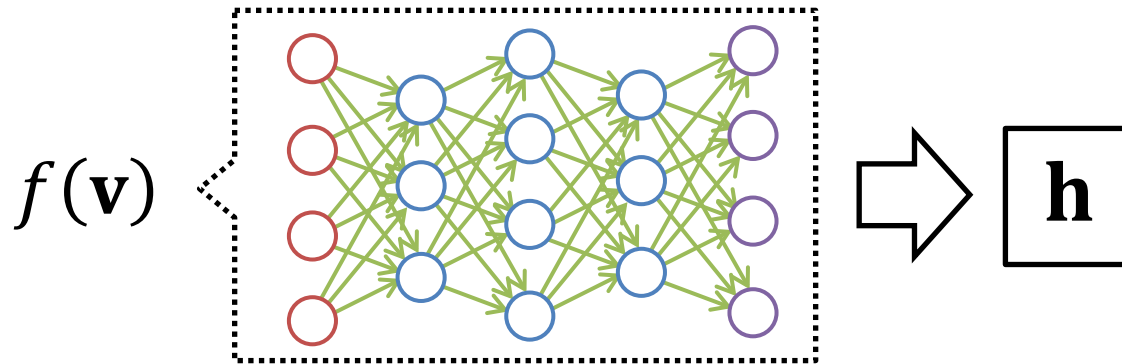
目次

1. 生成モデル
2. Restricted Boltzmann Machine
3. Deep Belief Network
4. 生成モデルの例

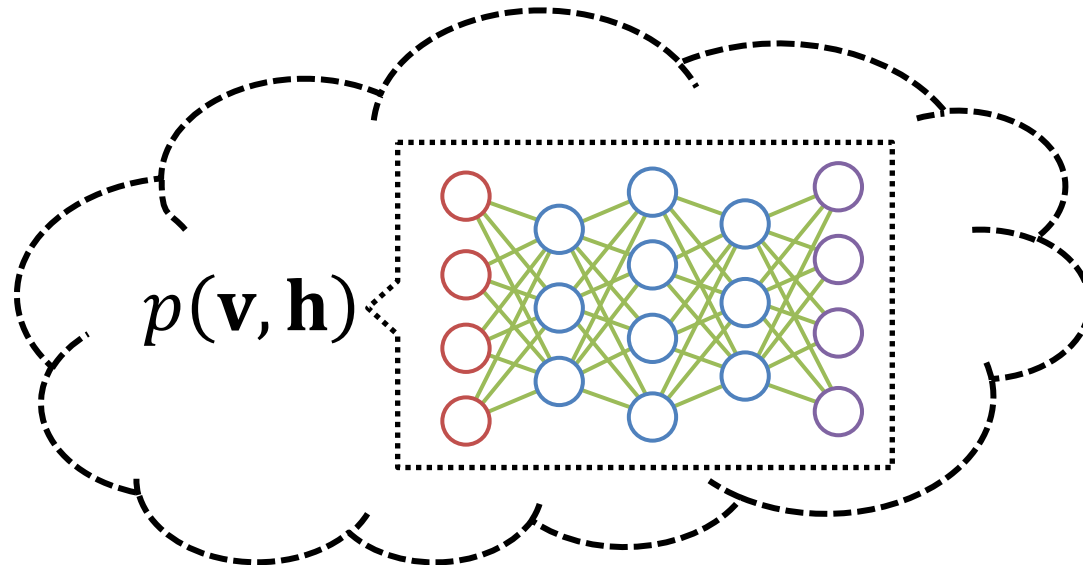
1. 生成モデル

生成モデルを用いたディープラーニング

これまで説明したディープラーニングの手法は、
ネットワークのふるまいは決定論的に記述される:

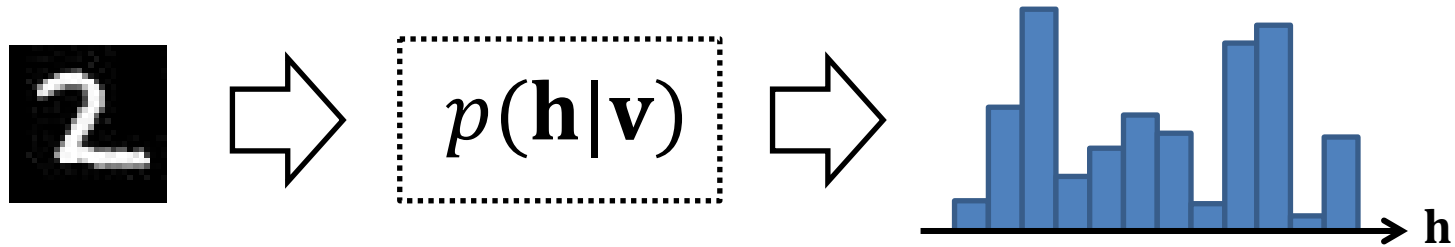


生成モデルを用いたディープラーニングの手法では、
ネットワークのふるまいを確率的にとらえる:

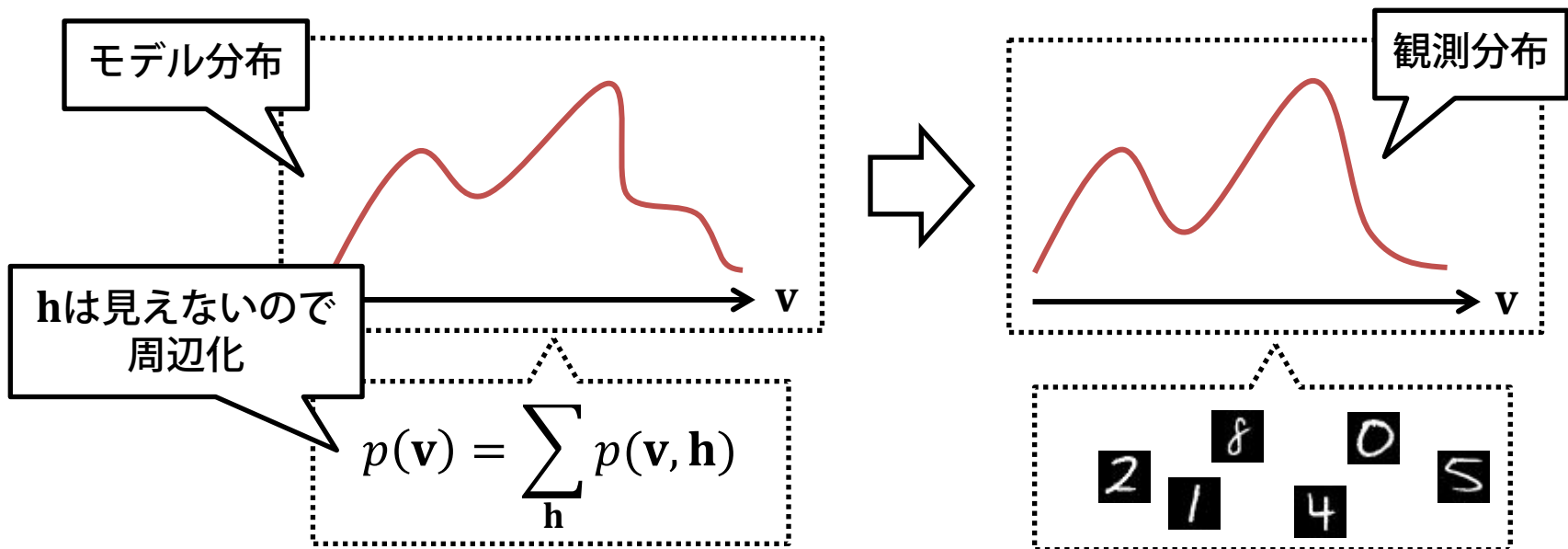


生成モデルの推定と学習

- 推定には、入力が与えられた際の隠れ層の条件付き確率を用いる

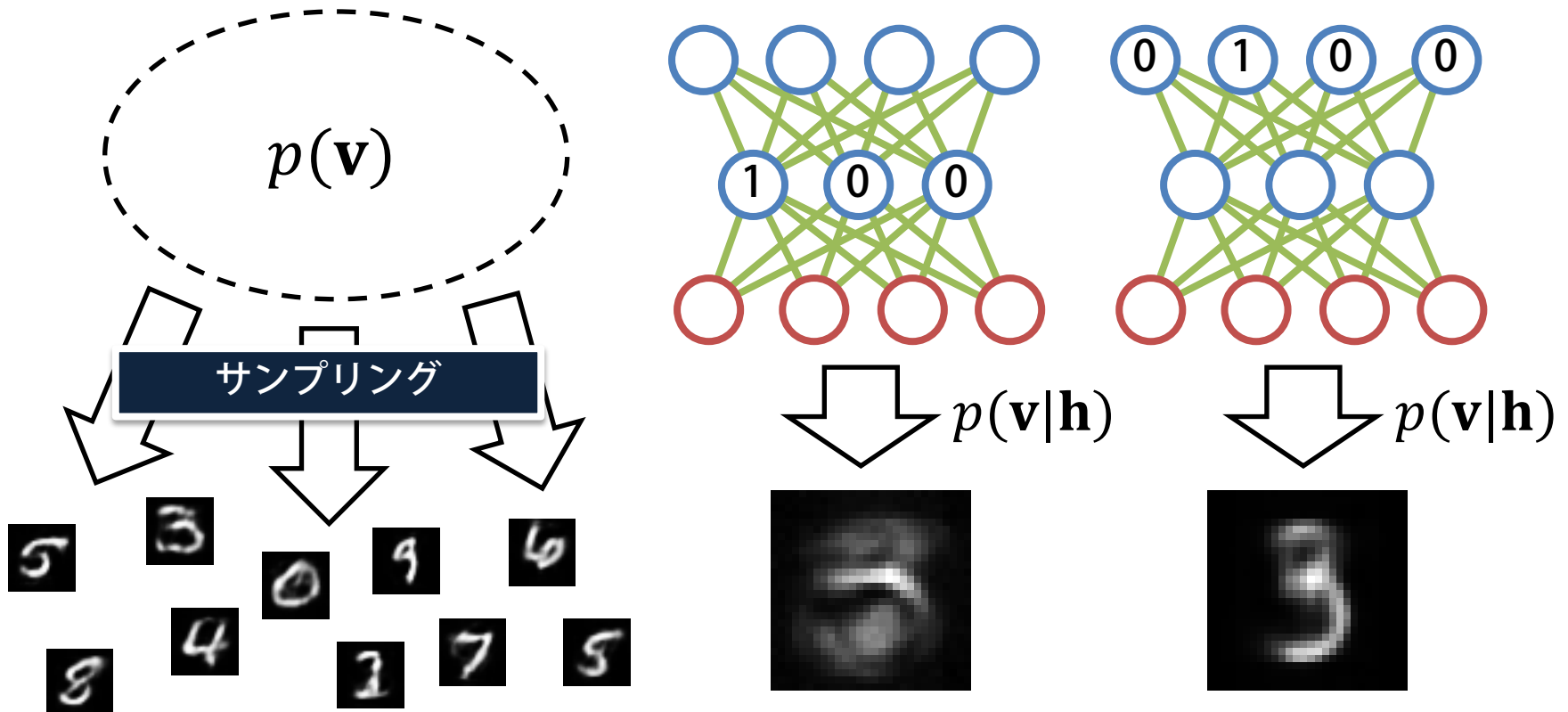


- パラメータの学習には、学習サンプルによって得られた観測分布に生成モデルを近づけることを行う



生成モデルの利点

1. 入力データそれ自体をモデルから生成できる
 - 隠れ層の学習結果を可視化しやすい
2. すべてを統計的な枠組みで議論できる
 - 推論やパラメータ学習などの各手法をシステムチックに求められる



エネルギー関数を用にしたモデリング

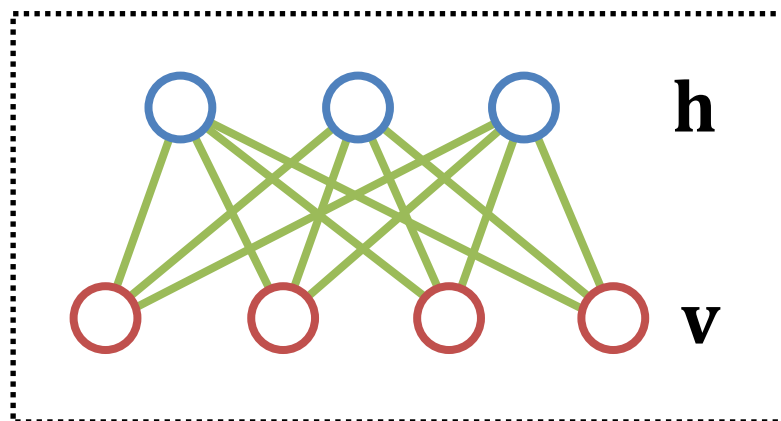
- 確率分布をエネルギー関数を用いてモデリング
- エネルギー関数によって同時確率変数が定まる構造

$$p(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{Z(\theta)} \exp(-E(\mathbf{v}, \mathbf{h}; \theta))$$

ボルツマン分布

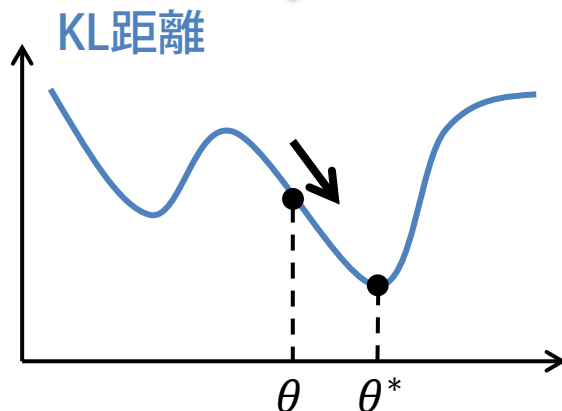
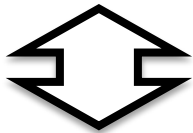
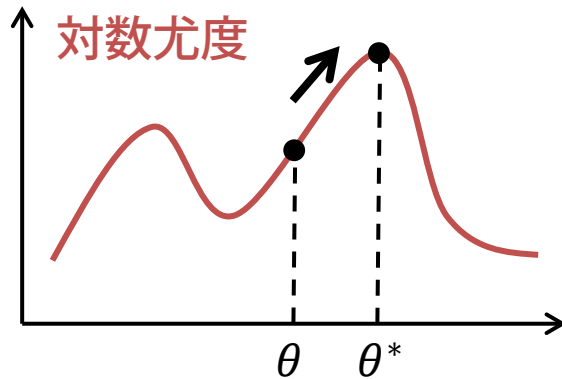
$$Z(\theta) = \sum_{\mathbf{v}} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}; \theta))$$

正規化定数
(分配関数)



$$E(\mathbf{v}, \mathbf{h}; \theta) \Rightarrow p(\mathbf{v}, \mathbf{h}; \theta)$$

尤度最大化による学習, KL距離最小化



- 対数尤度の最大化でパラメータを学習

$$\theta^* = \operatorname{argmax}_{\theta} \sum_n \ln p(\mathbf{v}^n; \theta)$$

- 対数尤度最大化はKL距離の最小化と等価
 - $p(\mathbf{v}; \theta)$ を観測分布 p_0 に近づける
 - KL距離: 確率分布間の近さを測る指標

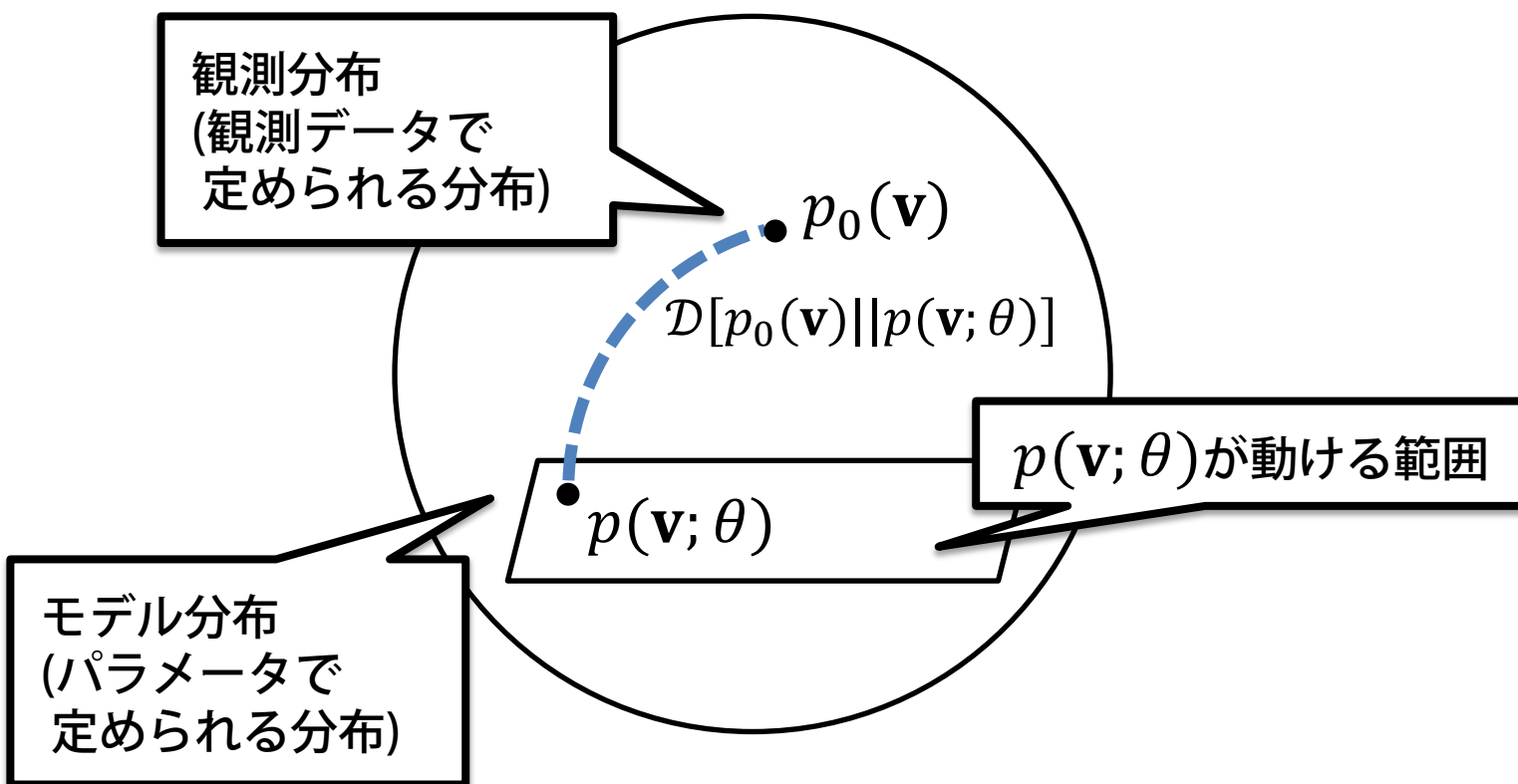
$$\theta^* = \operatorname{argmin}_{\theta} \mathcal{D}[p_0(\mathbf{v}) || p(\mathbf{v}; \theta)]$$

$$p_0(\mathbf{v}) = \frac{1}{N} \sum_n \delta(\mathbf{v} - \mathbf{v}^n)$$

観測分布
(観測データで定められる分布)

Kullback-Leibler(KL)距離

- KL距離: 確率分布間の近さを測る指標
- KL距離の最小化: モデル分布 p を観測分布 p_0 にできるだけ近づける



勾配法による最小化

- 勾配法によってパラメータを求める

$$\frac{\partial \mathcal{D}}{\partial \theta} = \left\langle \frac{\partial E}{\partial \theta} \right\rangle_{\text{data}} - \left\langle \frac{\partial E}{\partial \theta} \right\rangle_{\text{model}}$$

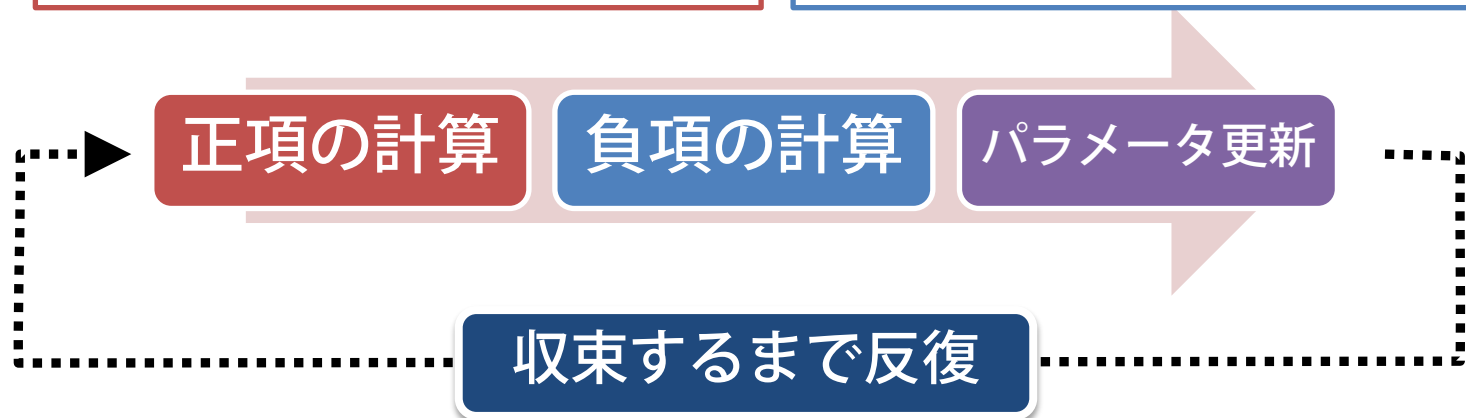
$\langle \cdot \rangle$ は期待値
をあらわす

正項(Positive Phase)

$$p_{\text{data}} = p_0(\mathbf{v})p(\mathbf{h}|\mathbf{v}; \theta)$$

負項(Negative Phase)

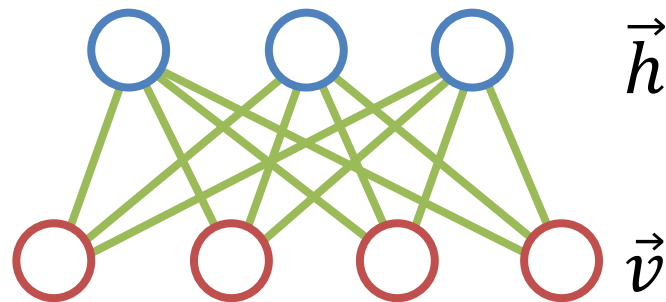
$$p_{\text{model}} = p(\mathbf{v}, \mathbf{h}; \theta)$$



2. RESTRICTED BOLTZMANN MACHINE

Restricted Boltzmann Machine(RBM)

- 最も単純なボルツマンマシンの一つ
- 多層でない。入力層同士, 隠れ層同士に結合が存在しない
- v_i, h_j はそれぞれ $\{0, 1\}$ の二値をとるものとする



$$E(\mathbf{v}, \mathbf{h}; \theta) = - \sum_i b_i v_i - \sum_j c_j h_j - \sum_{i,j} v_i W_{ij} h_j$$

バイアス

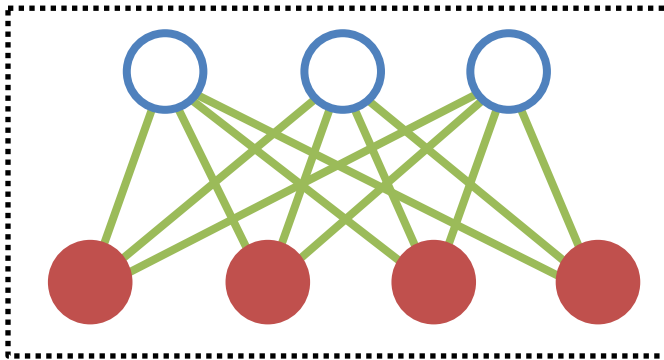
ウェイト

RBMの推論

😊 RBMの性質から $p(\mathbf{h}|\mathbf{v}; \theta)$, $p(\mathbf{v}|\mathbf{h}; \theta)$ は解析的に計算可能

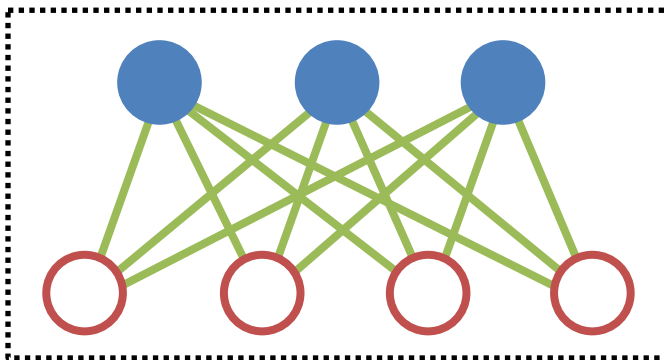


容易に計算できる



$$p(h_j = 1 | \mathbf{v}; \theta) = \sigma \left(c_j + \sum_j v_i W_{ij} \right)$$

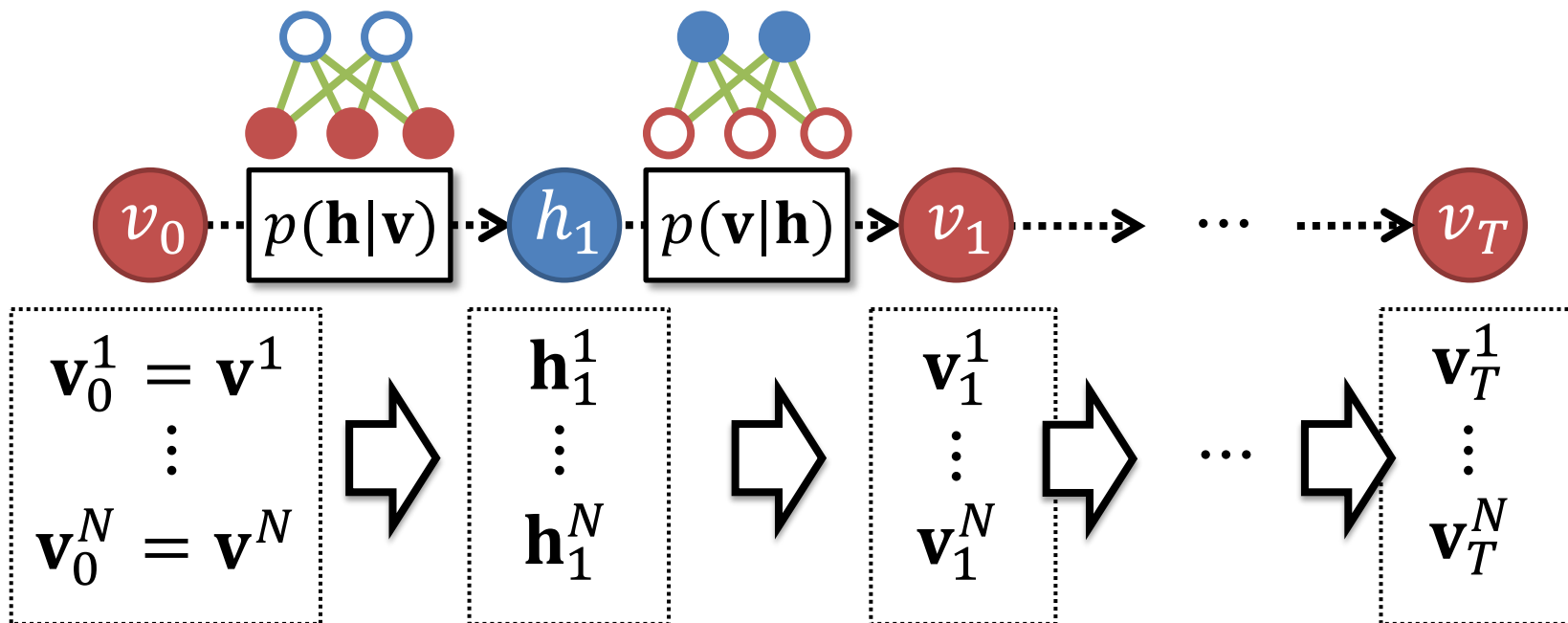
推論に使用



$$p(v_i = 1 | \mathbf{h}; \theta) = \sigma \left(b_i + \sum_j W_{ij} h_j \right)$$

RBMの学習(ギブスサンプリング)

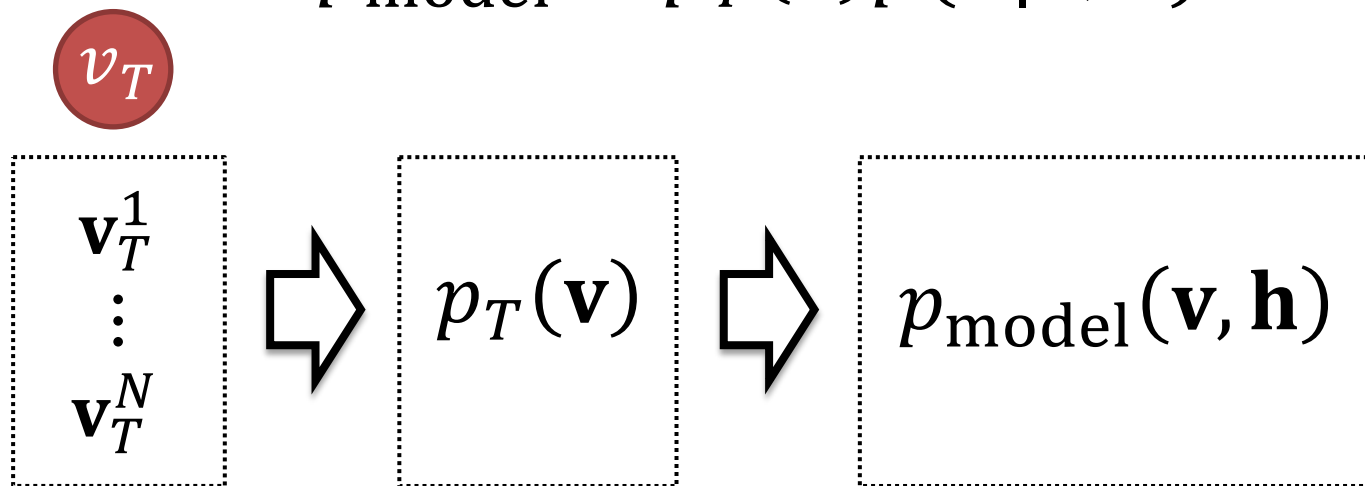
- 😊 正項に関しては, RBMの性質から解析的に計算可能
- ☹️ 負項に関しては, 計算コストが指数的に増大するため困難
- ギブスサンプリング法を用いて近似的に負項を求める



RBMの学習(ギブスサンプリング)

- $T \rightarrow \infty$ の極限において, \mathbf{v}_T は分布 $p(\mathbf{v}; \theta)$ に従う
 - $p(\mathbf{v}; \theta)$ に近い分布 $p_T(\mathbf{v})$ を生成, 近似的に負項を計算

$$p_{\text{model}} \approx p_T(\mathbf{v})p(\mathbf{h}|\mathbf{v}; \theta)$$



RBMの学習(Contrastive Divergence(Hinton, 2002))

T を有限回で打ち切った場合は？

⇒ **Contrastive Divergence**と呼ばれる目的関数の最小化に対応

$T \rightarrow \infty$

観測分布 モデル分布

$$\operatorname{argmin}_{\theta} \mathcal{D}[p_0(\mathbf{v}) || p(\mathbf{v}; \theta)]$$

$T \in \mathbb{N}$

T回のサンプリングで生成した分布 モデル分布

$$\operatorname{argmin}_{\theta} \mathcal{D}[p_0(\mathbf{v}) || p(\mathbf{v}; \theta)] - \mathcal{D}[p_T(\mathbf{v}) || p(\mathbf{v}; \theta)]$$

サンプル数が十分多く
 $T \rightarrow \infty$ のとき0となる

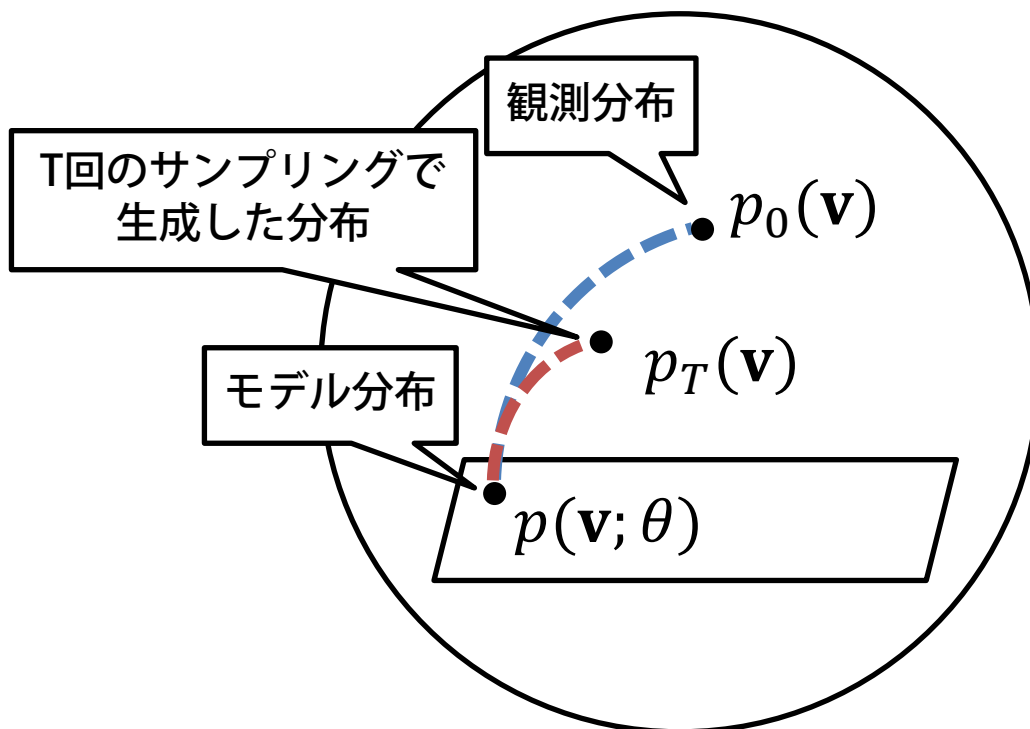
RBMの学習(Contrastive Divergence(Hinton, 2002))

T を有限回で打ち切った場合は？

⇒ **Contrastive Divergence**と呼ばれる目的関数の最小化に対応

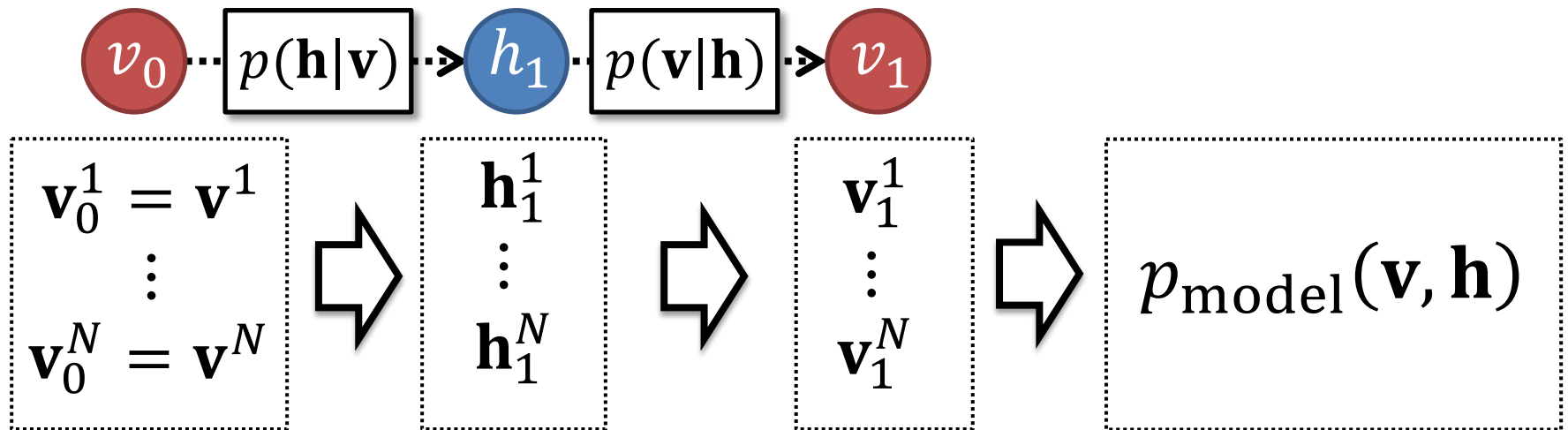
$T \in \mathbb{N}$

$$\operatorname{argmin}_{\theta} \mathcal{D}[p_0(\mathbf{v}) || p(\mathbf{v}; \theta)] - \mathcal{D}[p_T(\mathbf{v}) || p(\mathbf{v}; \theta)]$$



RBMの学習(Contrastive Divergence(Hinton, 2002))

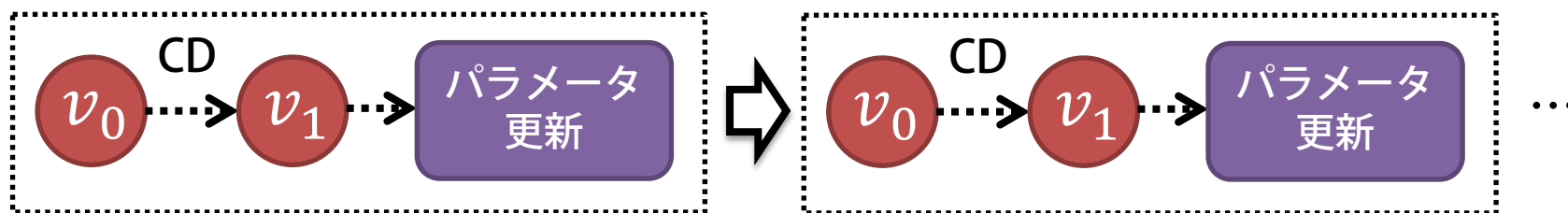
- T を有限回で打ち切った場合のパラメータの更新則は,
Contrastive Divergenceを目的関数とした最小化に対応する.
- **RBMは $T = 1$ でも精度のよい学習結果が得られる**



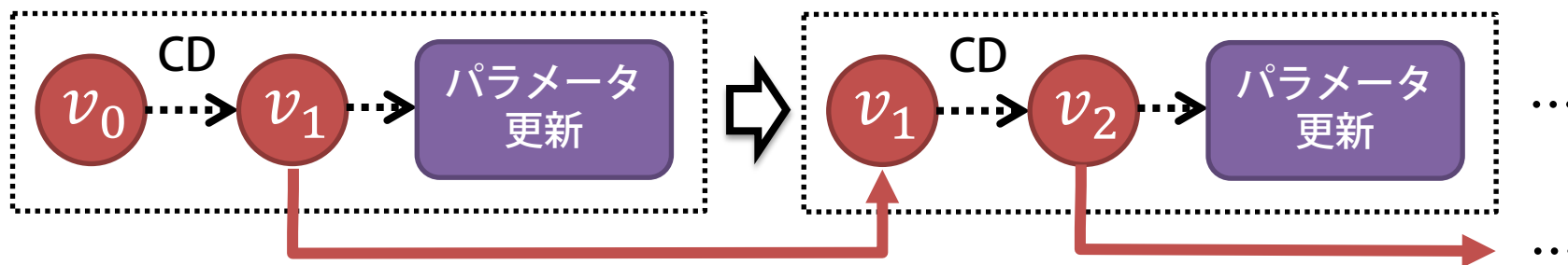
RBMの学習(負項の近似: Persistent CD(Tieleman, 2009))

- **Persistent CD**: Contrastive Divergenceを改良
- v_1 を次回のパラメータ更新の際に使用

CD法

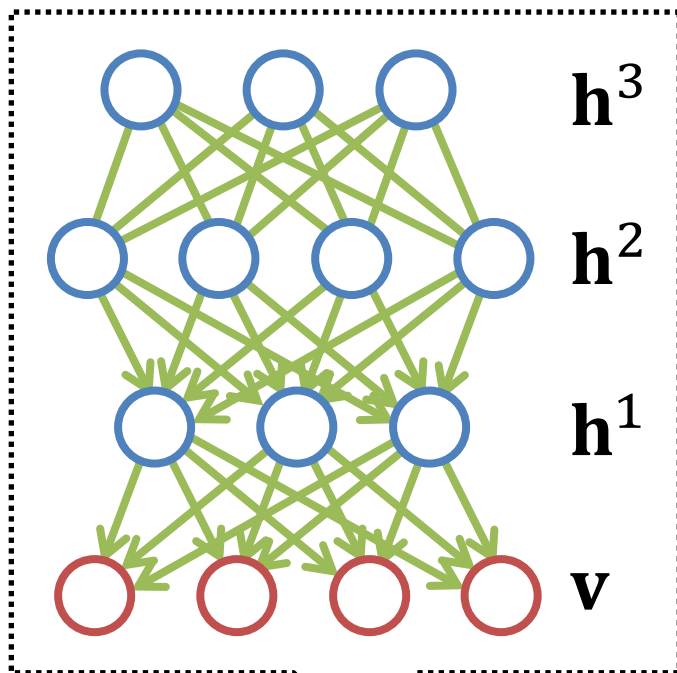


PCD法



3. DEEP BELIEF NETWORK

Deep Belief Network(Hinton, 2006)



- ディープラーニング流行のきっかけ
- 上層間だけ無向, それ以外は有向のグラフィカルモデル
- 学習を **プレトレーニング** と **微調整** の2段階に分ける

$$p(\mathbf{v}, \mathbf{h}; \theta) = p(\mathbf{v}|\mathbf{h}^1) \left(\prod_{l=1}^{L-2} p(\mathbf{h}^l|\mathbf{h}^{l+1}) \right) p(\mathbf{h}^{L-1}, \mathbf{h}^L)$$

$$\sigma \left(c_i^l + \sum_j W_{ij}^{l+1} h_j^{l+1} \right)$$

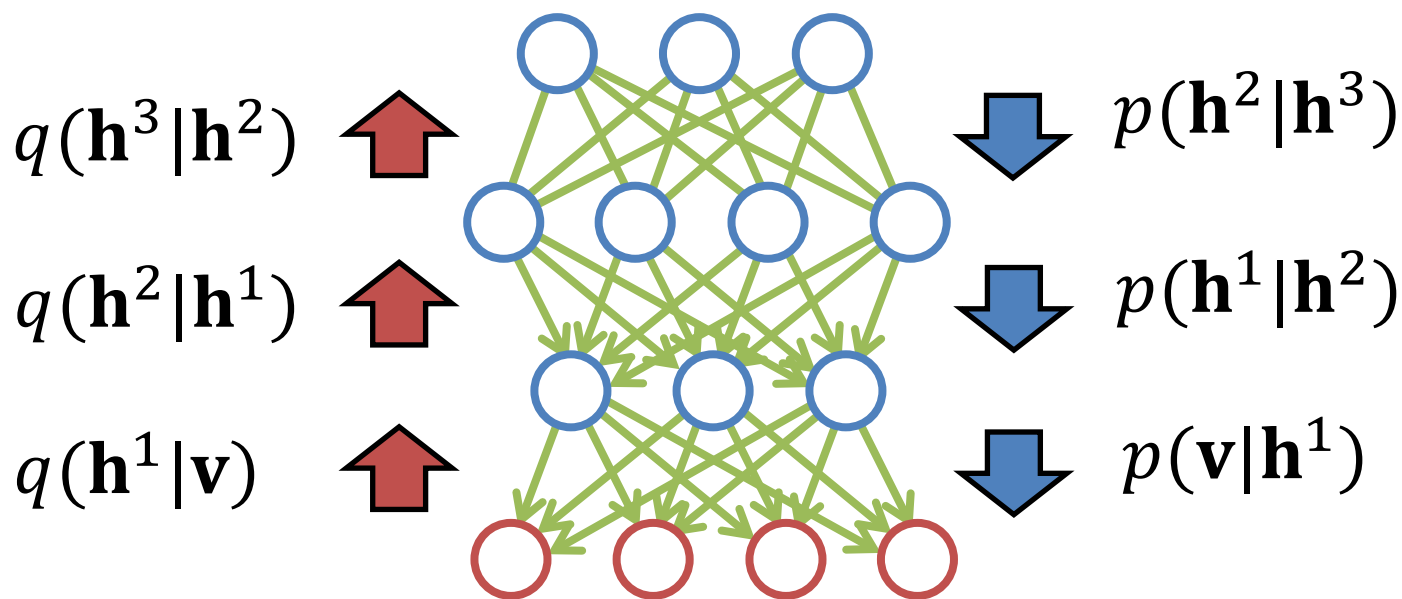
RBMの条件付き確率と同じ

RBMの結合確率密度と同じ

Deep Belief Networkの推論

☹️ 周辺分布 $p(\mathbf{h}^3 | \mathbf{v})$ は解析的に計算できない

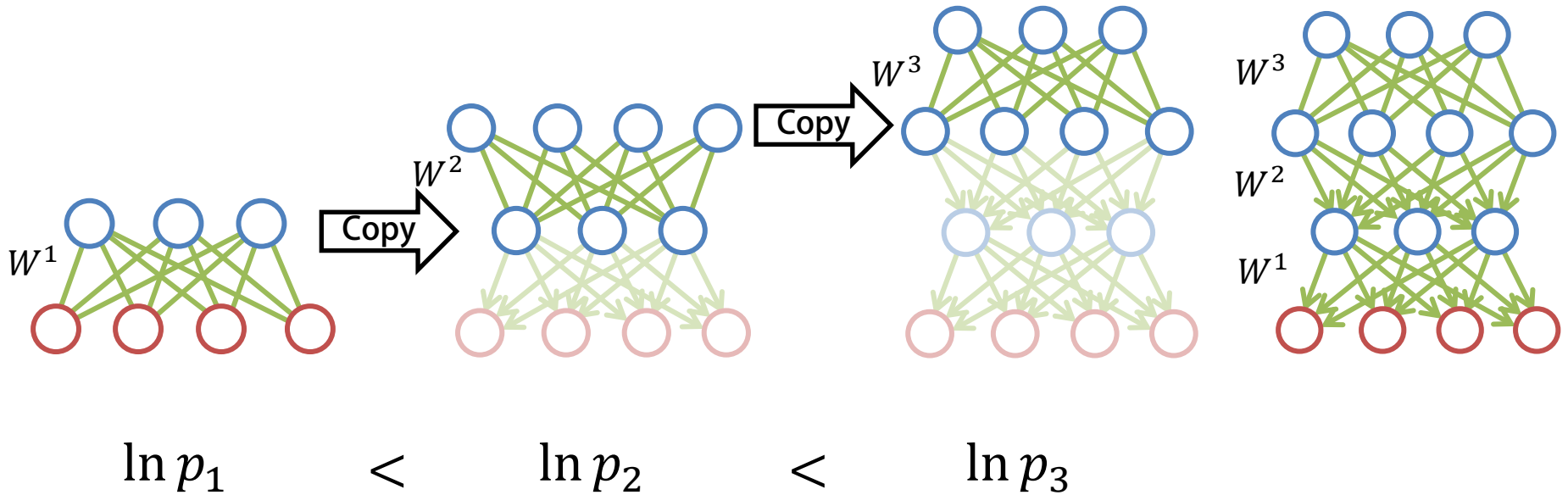
⇒ 近似分布 q を導入し, 入力層から隠れ層まで伝搬



$$q(h_j^l = 1 | \mathbf{h}^{l-1}) = \sigma \left(c_j^l + \sum_i h_i^{l-1} W_{ij}^l \right)$$

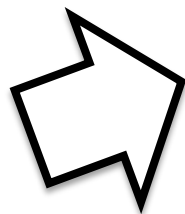
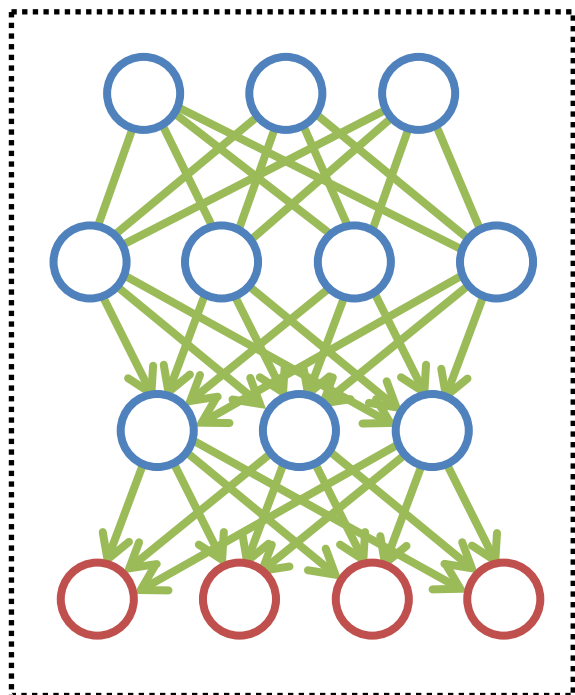
DBNの学習(プレトレーニング)

- RBMを用いて逐次的にトレーニングを行う
- プレトレーニングによる学習は, 対数尤度を必ず増加させる

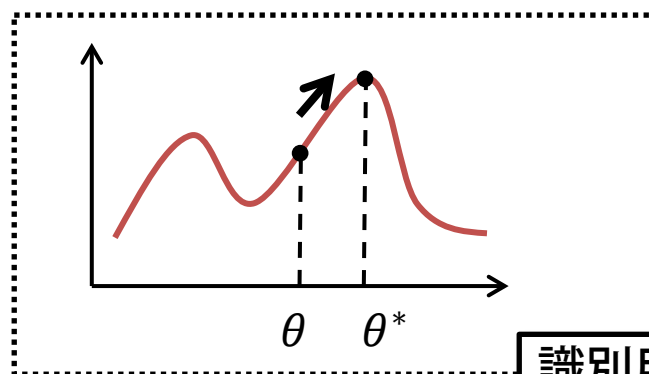


DBNの学習(微調整)

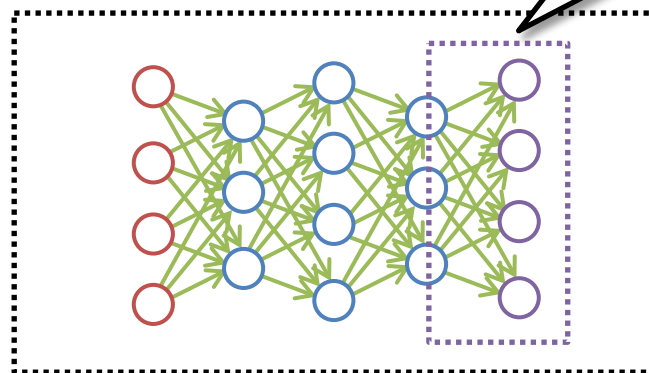
1. DBN全体の尤度を最大化するように調整
2. 決定論的な別の構造(Neural Net等)に写し, その中で教師付き学習を行う



尤度最大化



教師付き学習(NN)

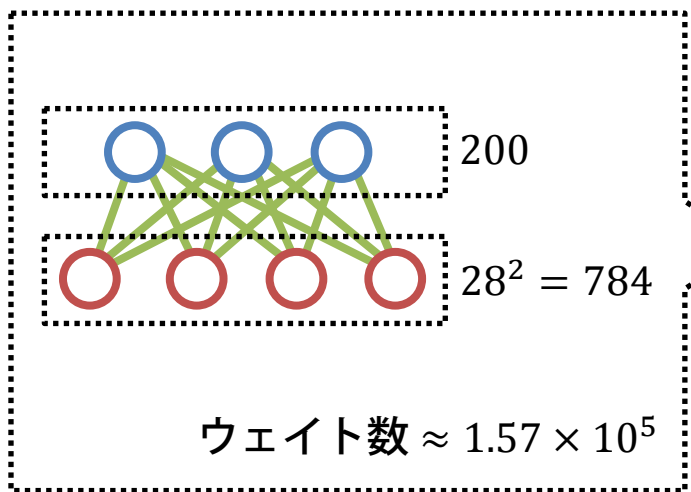


識別用の層を追加

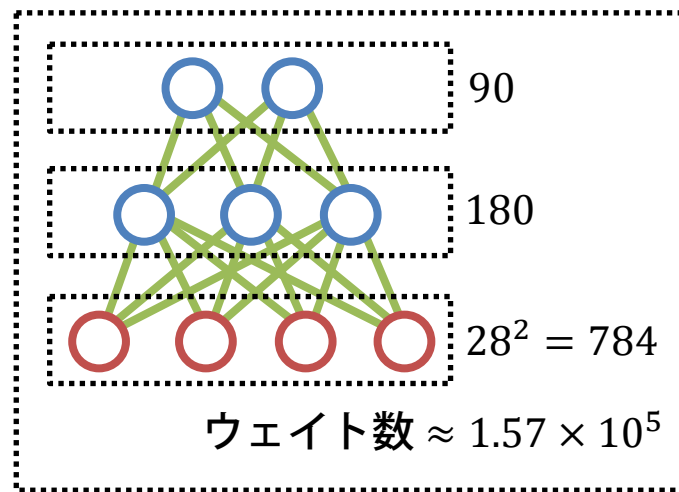
4. 生成モデルの例

Deep Boltzmann Machine(MNIST)

- 2層と3層のDBMモデルを, MNIST(手書き数字のデータセット)で学習
- ギブスサンプリングを用いて $p(\mathbf{v})$ のサンプルを生成



$p(\mathbf{v})$



$p(\mathbf{v})$

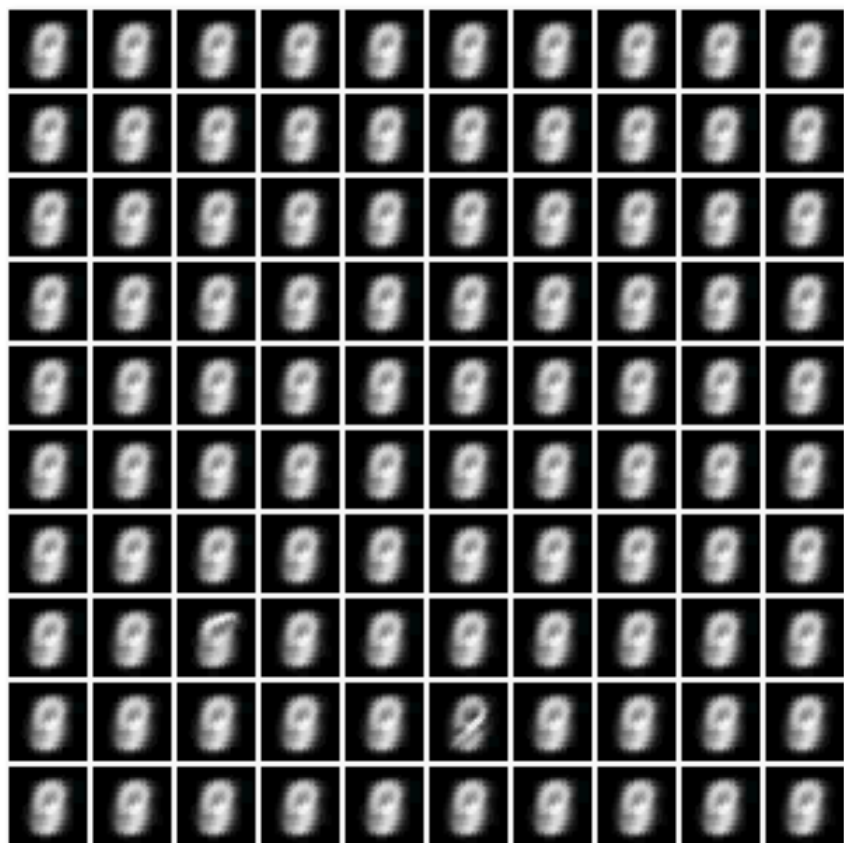
学習データ



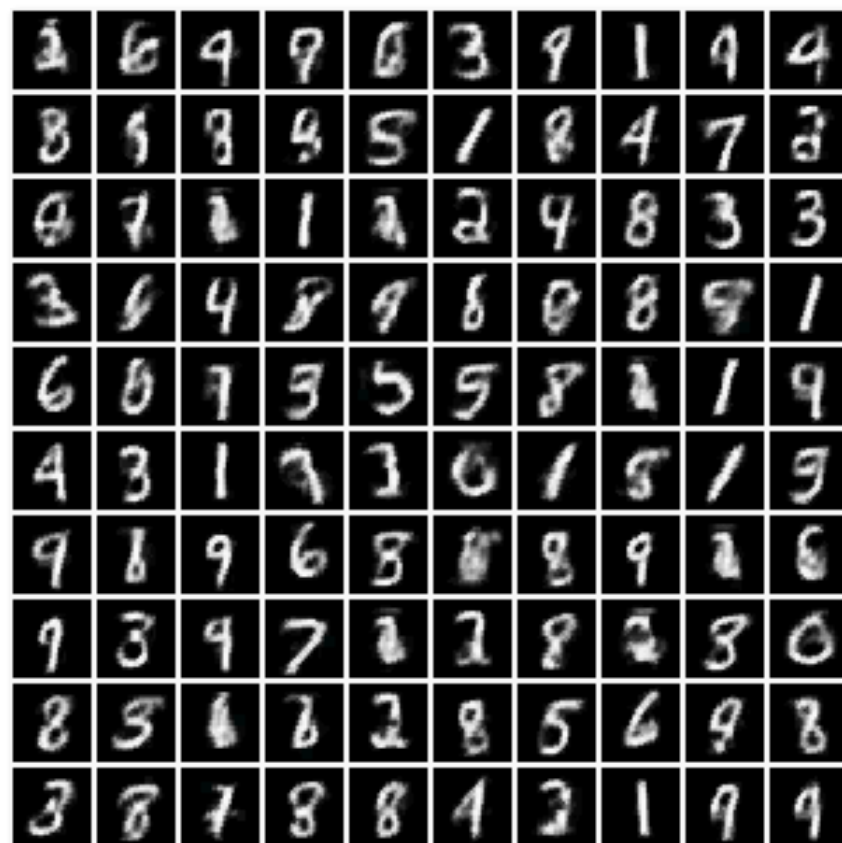
Deep Boltzmann Machine(MNIST)

- 2層と3層のDBMモデルを, MNIST(手書き数字のデータセット)で学習
- ギブスサンプリングを用いて $p(v)$ のサンプルを生成

2層

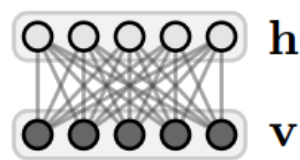
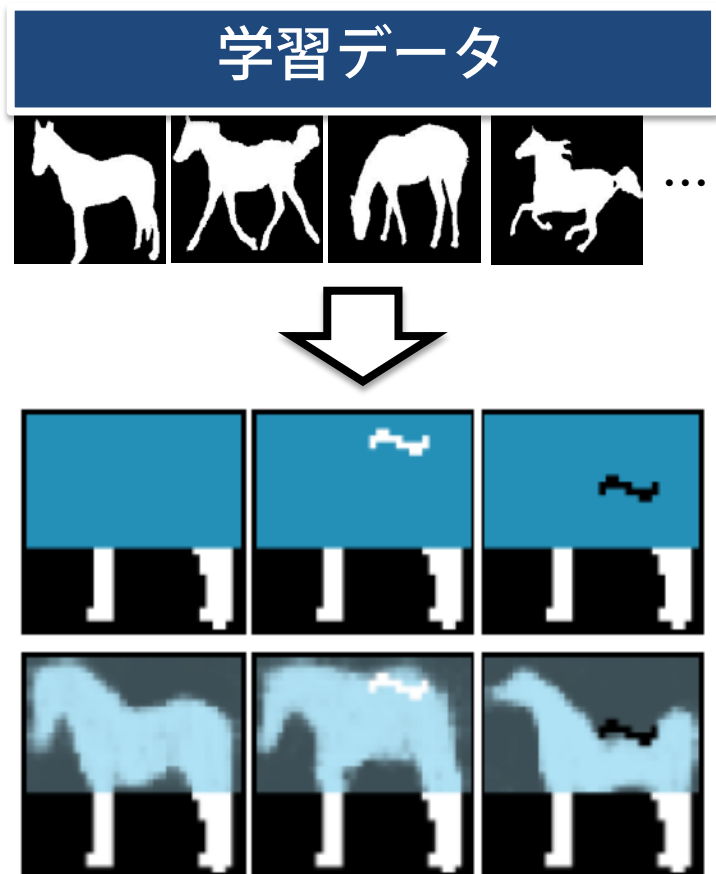


3層

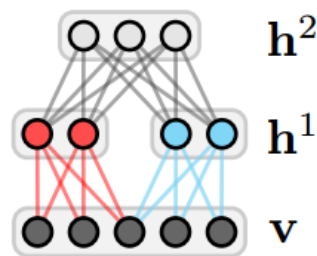


Shape Boltzmann Machine(Eslami et. al., 2012)

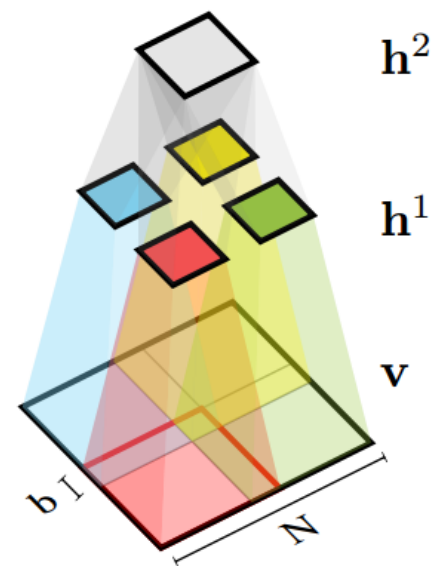
- 隠された領域を復元する生成モデル。ボルツマンマシンを使用
- 汎化能力に優れる
- パラメータを削減するため、局所的に結合した生成モデルを提案



(b) RBM



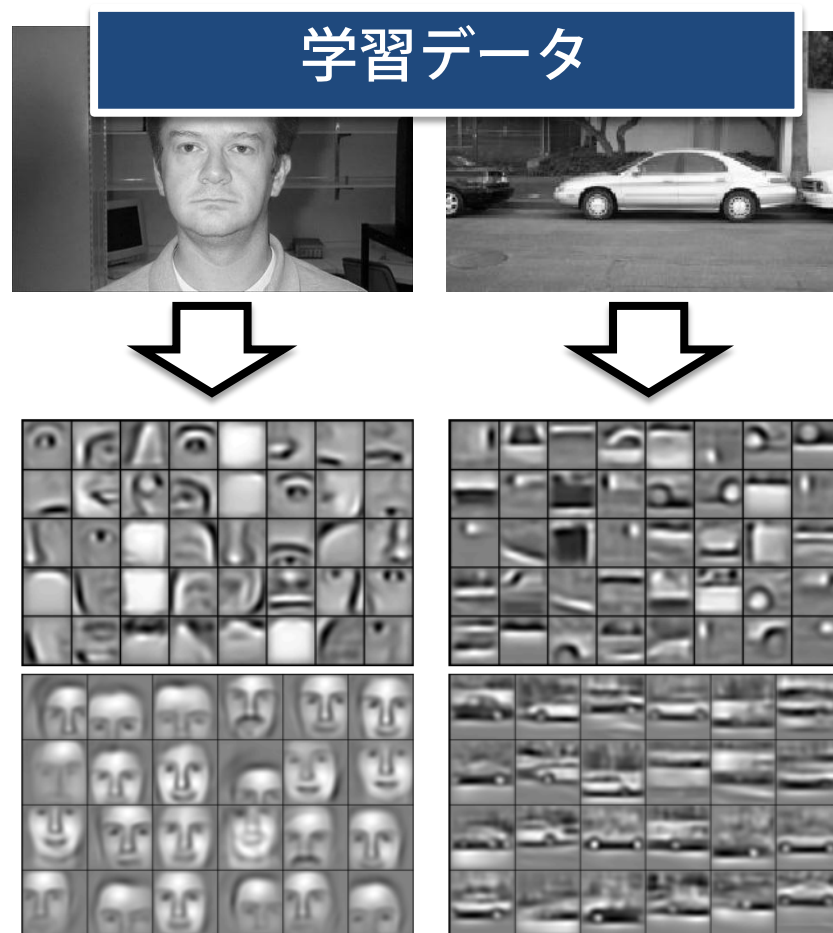
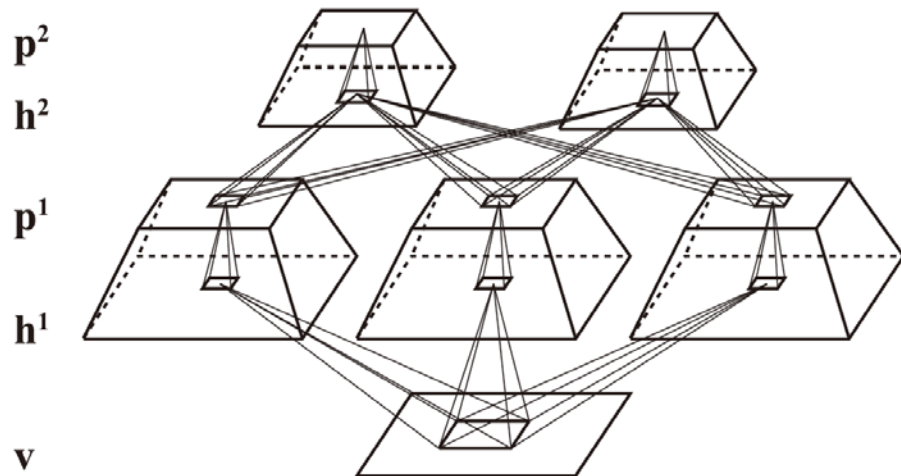
(d) ShapeBM



(e) ShapeBM

Convolutional DBN(Lee et. al. 2009)

- Convolutional NNのモデルを生成モデルに対して応用
- max-poolingを組み込む目的で, **probabilistic max-pooling**を提案




目次

- インTRODクシヨン
 - DNNの事例紹介
- NNの基礎と研究の略史
 - なぜNN研究は廃れたか？
- プレトレーニング
 - Encoder-decoder paradigm
- Convolutional Neural Networks (CNN)
 - プーリングと不変特徴
 - CNN vs. fully-connected NN
- ニューロサイエンス
 - どこまで分かっている？
- 生成モデル
 - Restricted Boltzmann Machine (RBM)
 - Deep Belief Network (DBN)
- 実践編
 - **cuda-convnet を使ったMNISTの学習**
- まとめ

cuda-convnet

<http://code.google.com/p/cuda-convnet/>


- ILSVRC2012チャンピオンのSupervisionチームのコード
- nVidia GTX580系統のGPU
- Linux と CUDA

 **cuda-convnet**
High-performance C++/CUDA implementation of convolutional neural networks

[Project Home](#) [Wiki](#) [Issues](#) [Source](#)

Summary [People](#)

Project Information

 +10 Recommend this on Google

★ Starred by 56 users
[Project feeds](#)

Code license
[New BSD License](#)

Labels
Machinelearning, Academic, Cuda, CPlusPlus, Python, neuralnet, convnet, convolution, GPU

Members
[akrizhev_@gmail.com](#)

This is a fast C++/CUDA implementation of convolutional (or more generally, feed-forward) neural networks. It can model arbitrary layer connectivity and network depth. Any directed acyclic graph of layers will do. Training is done using the back-propagation algorithm.

Fermi-generation GPU (GTX 4xx, GTX 5xx, or Tesla equivalent) required.

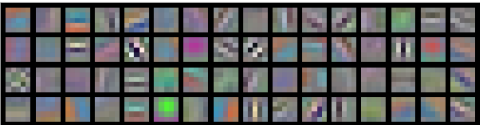
Documentation

- [Compiling](#) -- how to check out and compile this code.
- [Data](#) -- what kind of data this net can train on.
- [LayerParams](#) -- how to specify an architecture for the net.
- [NeuronTypes](#) -- types of hidden unit nonlinearities.
- [TrainingNet](#) -- how to train the net.
- [Options](#) -- the command-line arguments that the net takes.
- [ViewingNet](#) -- how to look inside the checkpoints saved by the net.
- [CheckingGradients](#) -- how to numerically test the gradients for correctness.

Fast results

- 11% error on [CIFAR-10](#) in 75 minutes, with image translations and horizontal reflections ([def. params](#)).
- 13% error on CIFAR-10 in 25 minutes, with image translations and horizontal reflections ([def. params](#)).
 - See [Methodology](#) for details of training.

Filters learned by this net:



• 18% error on CIFAR-10 in 20 minutes, without any image translations/transformations/preprocessing ([def. params](#)).

• 26% error on CIFAR-10 in 80 seconds, without any image translations/transformations/preprocessing ([def. params](#)).

Recent changes

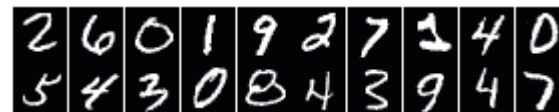
- Jul 17, 2012
 - Fixed bug in contrast normalization backpropagation code which caused wrong gradients to be computed near image borders. (Thanks [Hannes Schulz](#).)

cuda-convnet

<http://code.google.com/p/cuda-convnet/>

1. データファイル作成
 - Python のリスト（サンプルはランダムに並べる）を作り，cPickle で複数ファイル（batch と呼ぶ）に分割してセーブ
 - 複数個の訓練用 batch と 1 つ以上のテスト用 batch
2. データ読み込み関数を書く
 - データサイズとプレーン数，正規化の有無
3. アーキテクチャを考える
4. 学習の実行
 - 訓練用バッチのうち 1 つを validation 用
 - テスト誤差が減少し続けるまで，反復学習（数 10 ～ 数 100 エポック）
 - 全バッチを使って反復学習
 - 学習係数を小さくし，複数回（10 エポック程度）反復学習
 - 終了

MNIST

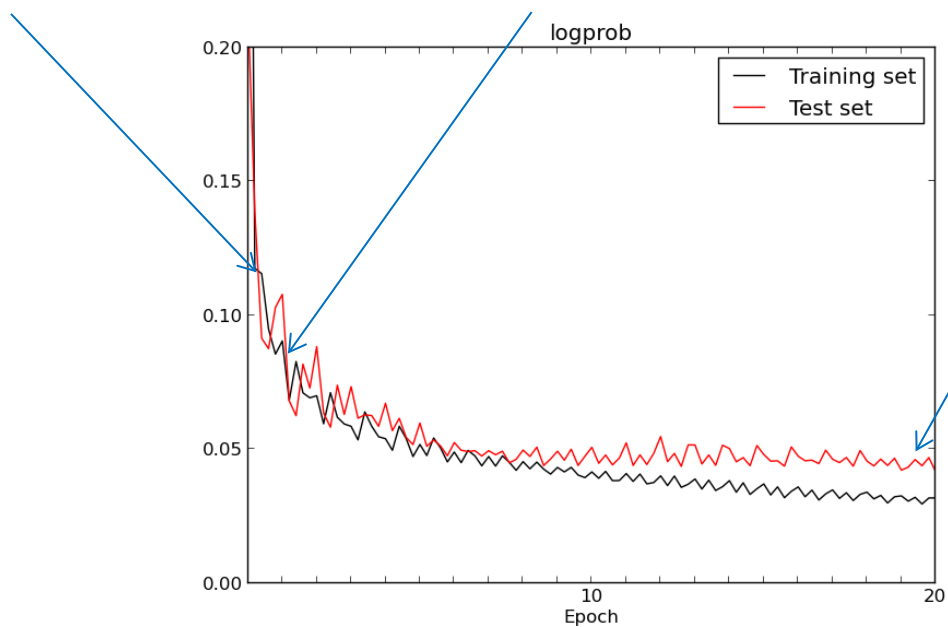
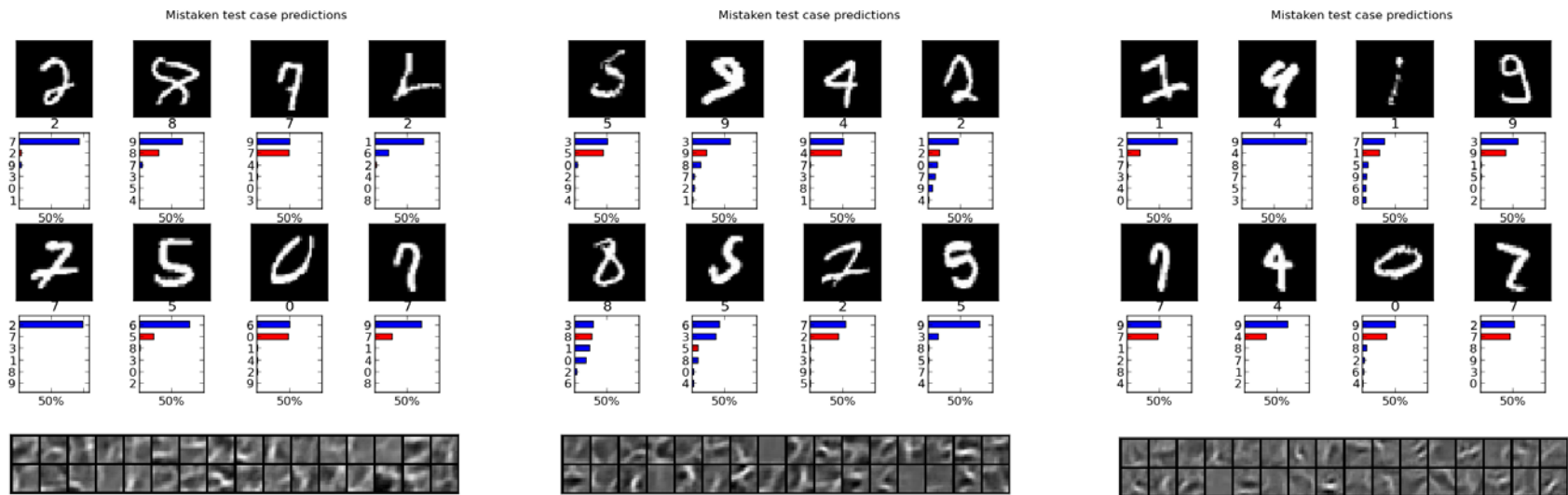


グレーレベル 28 × 28

60000 訓練・10000 テスト
サンプル → 訓練用 6 batch
テスト用 1 batch

cuda-convnet

<http://code.google.com/p/cuda-convnet/>



誤答率1.1%

Best result: 0.3%
(>人間)

目次

- インTRODクシヨン
 - DNNの事例紹介
- NNの基礎と研究の略史
 - なぜNN研究は廃れたか？
- プレトレーニング
 - Encoder-decoder paradigm
- Convolutional Neural Networks (CNN)
 - プーリングと不変特徴
 - CNN vs. fully-connected NN
- ニューロサイエンス
 - どこまで分かっている？
- 生成モデル
 - Restricted Boltzmann Machine (RBM)
 - Deep Belief Network (DBN)
- 実践編
 - cuda-convnet を使ったMNISTの学習
- まとめ

まとめ

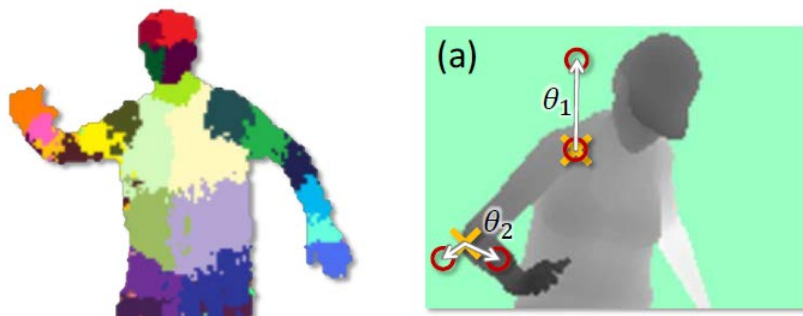
- DNNは大きな成果を上げている
 - 既存手法の性能比で大きなマージン
 - まだ性能向上の余地
- 特に大きな理論的進展があったわけではない（今となっては）
 - プレトレーニングは必要か？
 - CNNは80年代後半からほとんど変わらず
 - 「やればできる」ことがわかったこと・計算機の性能向上
- 性能を引き出すためのノウハウが必要なのは昔と変わらない
- エンジニアリングの勝負になるかもしれない
 - 大規模NNによる大規模データを使った学習
 - 認識性能上げたい → NN多層化 → 過学習のリスク → 学習データ量増やす → 計算性能必要
- Representation/feature learning への期待
 - semi-supervised / transfer / self-taught learning

“Representation(Feature) learning”

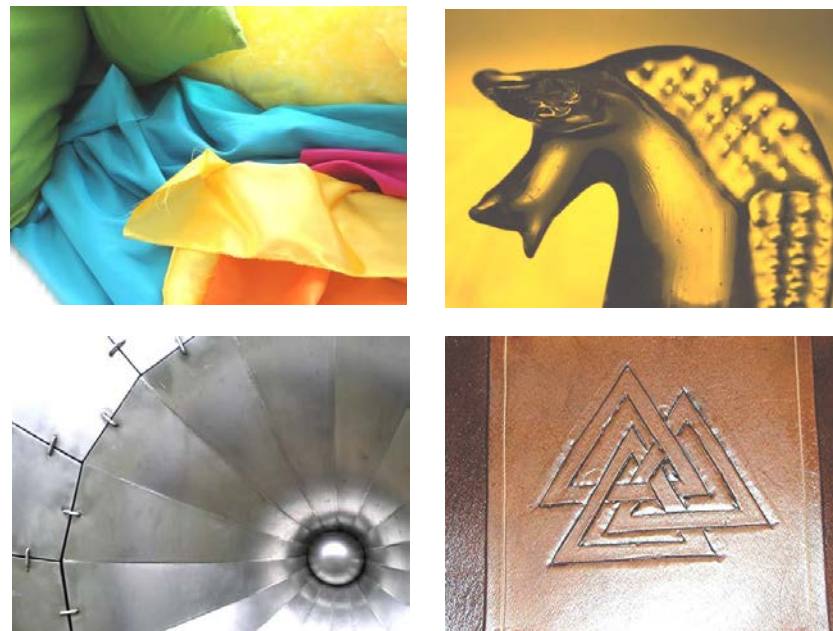
Bengio, Representation Learning: A Review and New Perspective, arXiv, Oct 2012

- 特徴抽出器を hand-craft or 学習で見つける？
- 複雑な問題 → 何を取り出したらいいかわからない

Kinect Mocap



マテリアル分類, 質感の画像認識



Schotton et al., Real-Time Human Pose Recognition in Parts from a Single Depth Image, CVPR2011

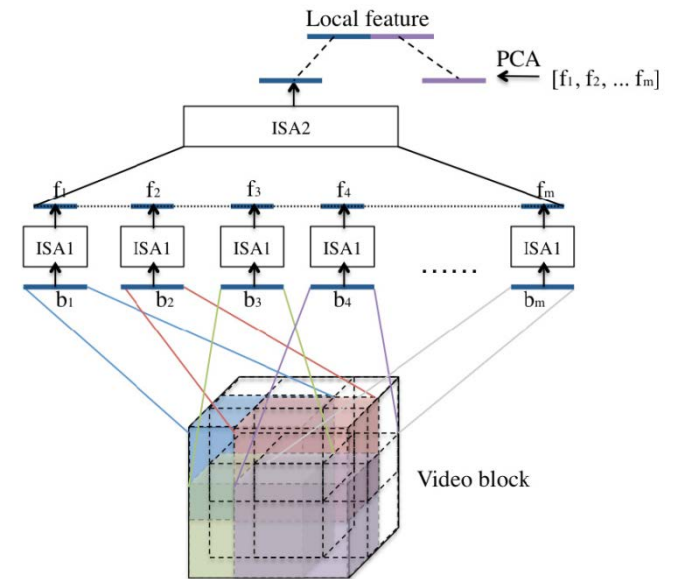
Liu et al., Exploring Features in a Bayesian Framework for Material Recognition, CVPR2010

“Representation(Feature) learning”

Bengio, Representation Learning: A Review and New Perspective, arXiv, Oct 2012

- 特徴抽出器を hand-craft する or 学習で見つける？
- 複雑な問題 → 何を取り出したらいいかわからない

人の行動認識



Le et al., Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis, CVPR2011