

SAPT

第4回知能化分科会

一色 浩

(SAPT知能化分科会長)

2018.06.09

SAPT電子会議室C

目次

1. 第2回分科会での話題: ニューロによる独立成分分析
2. 瀧 雅人著「これならわかる深層学習」
3. 講義担当者のニューロとの関わり
4. 飲み会の代わり...自由討論

第2回分科会での話題: ニューロによる独立成分分析

ZDNet Japan > 経営



理研、並列計算で感覚情報を分解するアルゴリズムを開発--“脳型コンピュータ”実現

へ一歩

<https://japan.zdnet.com/article/35085719/>

NO BUDGET 2016年07月16日 07時16分

いいね! 6

G+

B! 0

Pocket 26

印刷

メール ▼

ダウンロード ▼

クリップ

PR | 生産性向上は企業の至上命題--明日の働き方を考え、デジタルで未来を切り開く

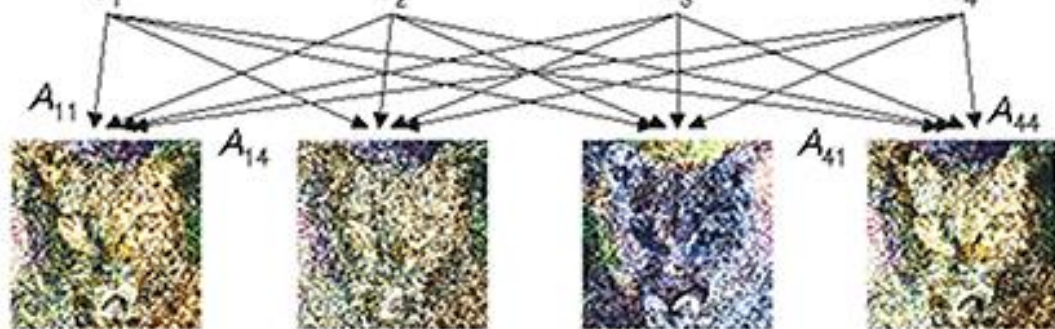
PR | 「頭おかしいぞ!」はイノベーションの瞬間かも? 記事まとめを読んで考えよう!

理化学研究所（理研）は6月29日、脳科学総合研究センター神経適応理論研究チームの豊泉太郎チームリーダーらの研究チームが、人の脳を模した“脳型コンピュータ”とも言える神経回路型ハードウェア（神経細胞を模倣した計算素子が互いに結合したネットワークからなる計算装置）を用いて複数の感覚入力を独立した成分に分解するためのアルゴリズムを開発したと発表した。

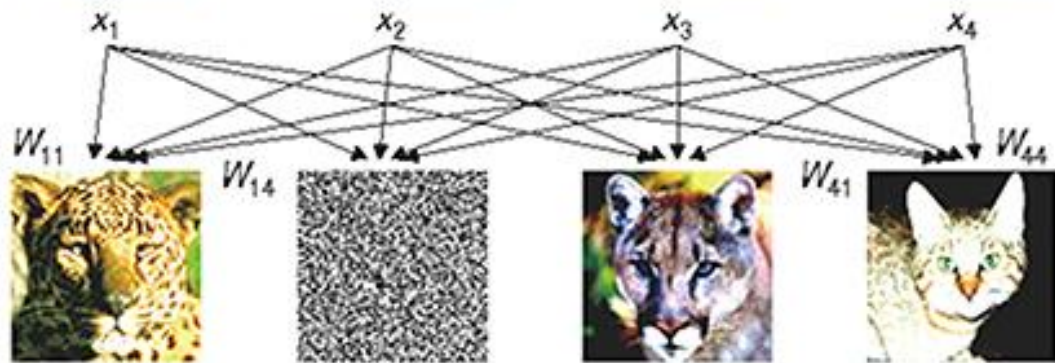
信号源



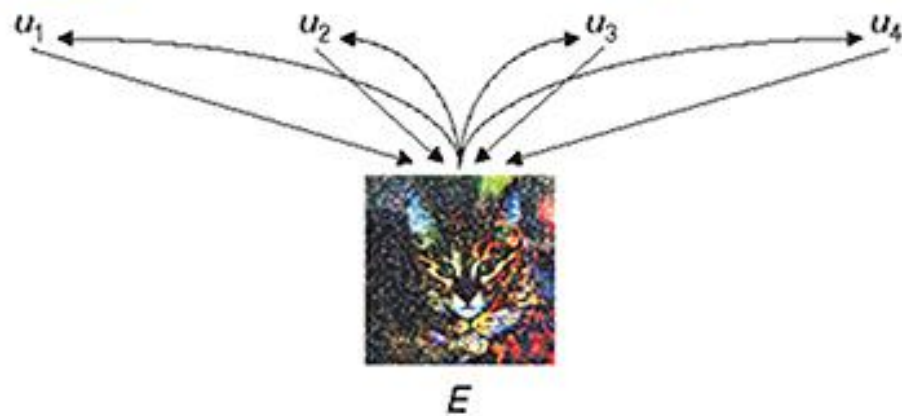
入力

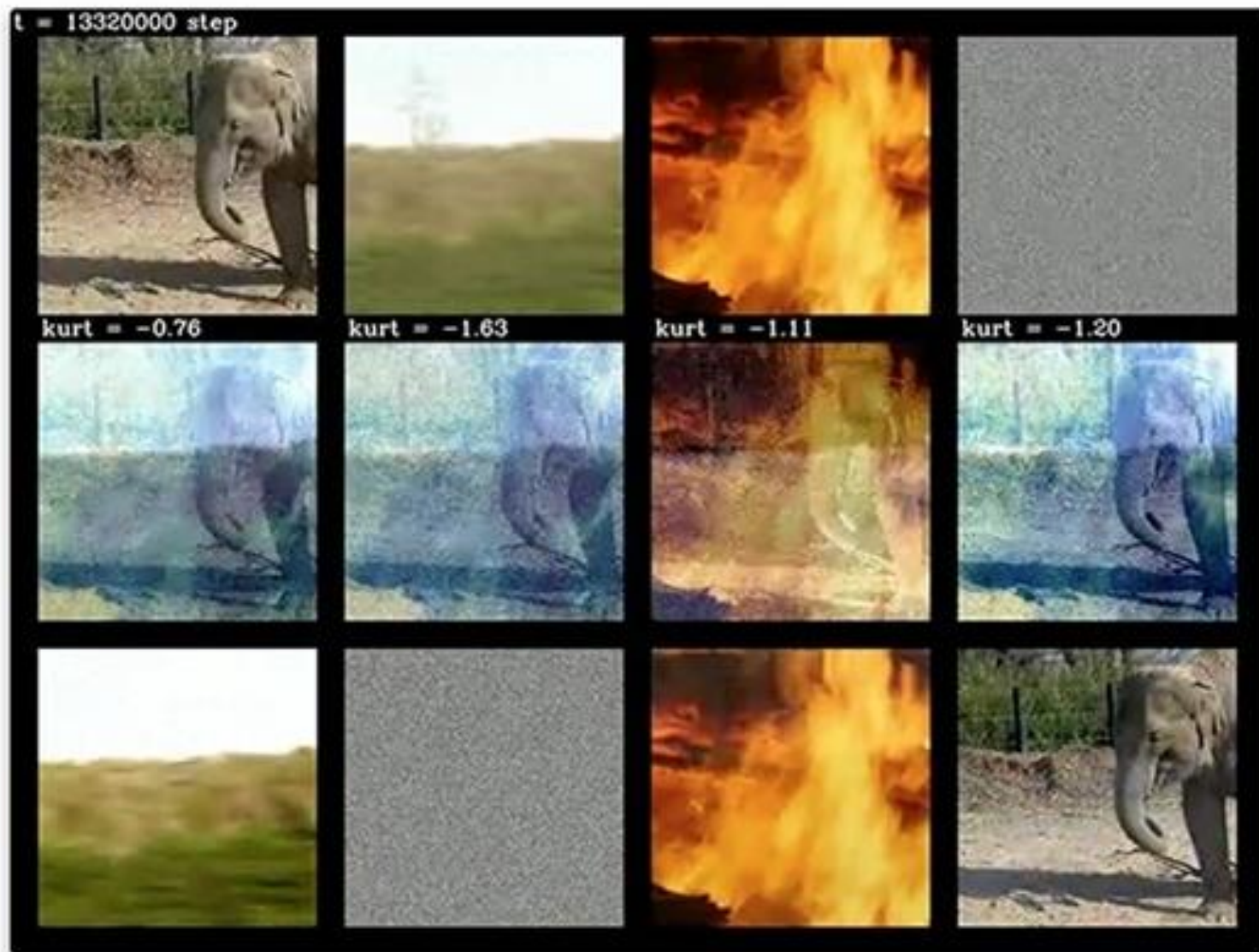


出力



広域信号





EGHRによる自然動画の独立成分への分解。信号源として4つの動画を用意し（上段）、それらを混ぜ合わせて入力を生成（中段）する。EGHRは自然動画に対してもICAを実行可能であり、出力（下段）は元の動画（信号源）をうまく復元することができたという

SCIENTIFIC REPORTS

OPEN

A Local Learning Rule for Independent Component Analysis

Takuya Isomura^{1,2,3} & Taro Toyoizumi¹

<http://www.nature.com/scientificreports/>

17ページ

人間は重なった信号を別々に認識できる。数学的にはICA(独立成分分析という)。簡単に実装できて信頼性の高い方法を提案する。他のローカル学習ルールよりも格段に性能が良い。本方法では前処理は不要である。

原データの確率分布を仮定し、最尤推定で分離

誤差評価には最尤推定(後出)を用いる？
最急降下法と解釈してもよい。

Error Gated Hebbian Rule (EGHR) というローカル学習ルールを提案する.

$$\tau_w \dot{W} = \langle (E_0 - E(\mathbf{u})) g(\mathbf{u}) \mathbf{x}^T \rangle$$

ここで $g(u_i)x_j$ はstandard Hebbian term

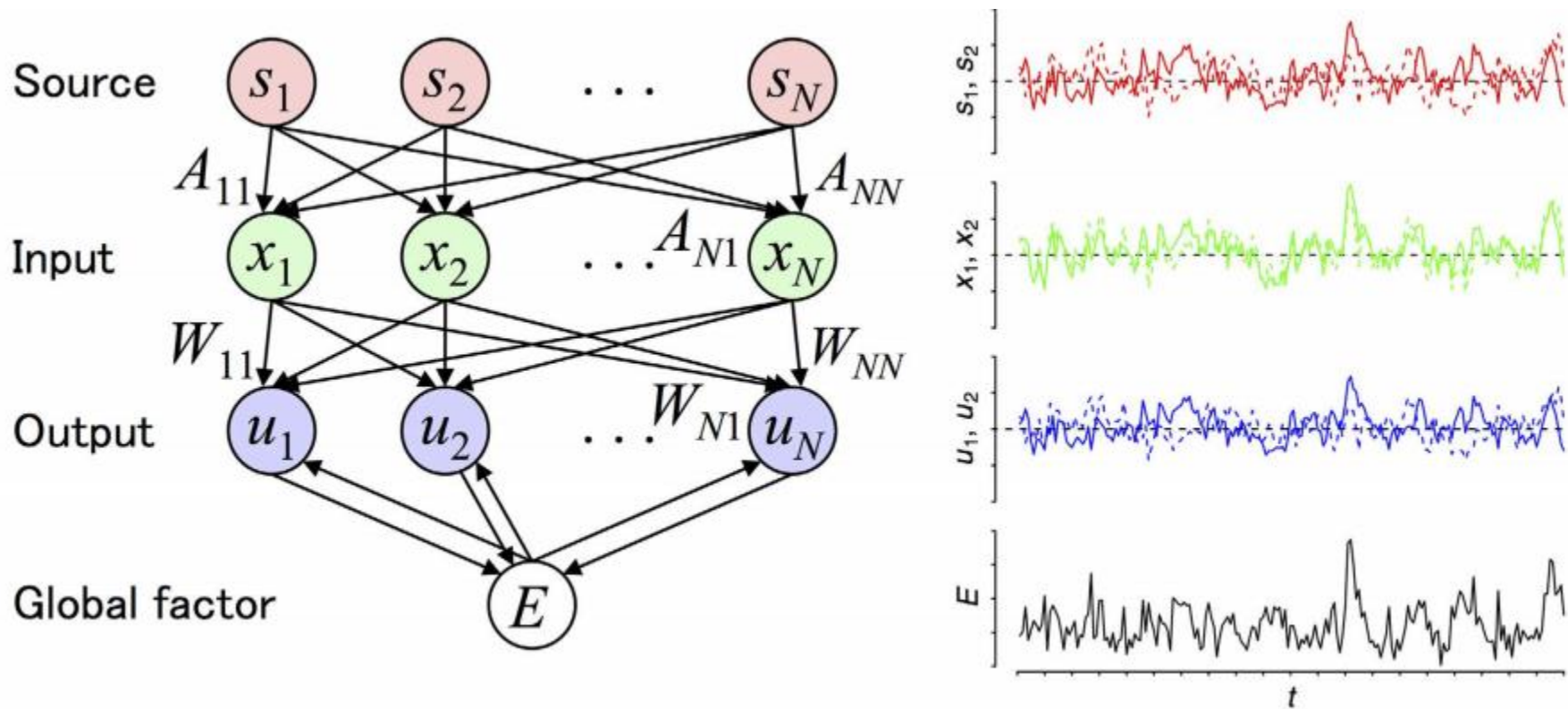
$g(u_i) = -d \log p(u_i) / du_i$ はpost synaptic factor

x_j はpre synaptic factor

$\langle \bullet \rangle$ はexpectation over ensemble \mathbf{x}

EGHR は, $\langle (E_0 - E(\mathbf{u}))^2 / 2 \rangle$ をコスト関数とする
最急降下法

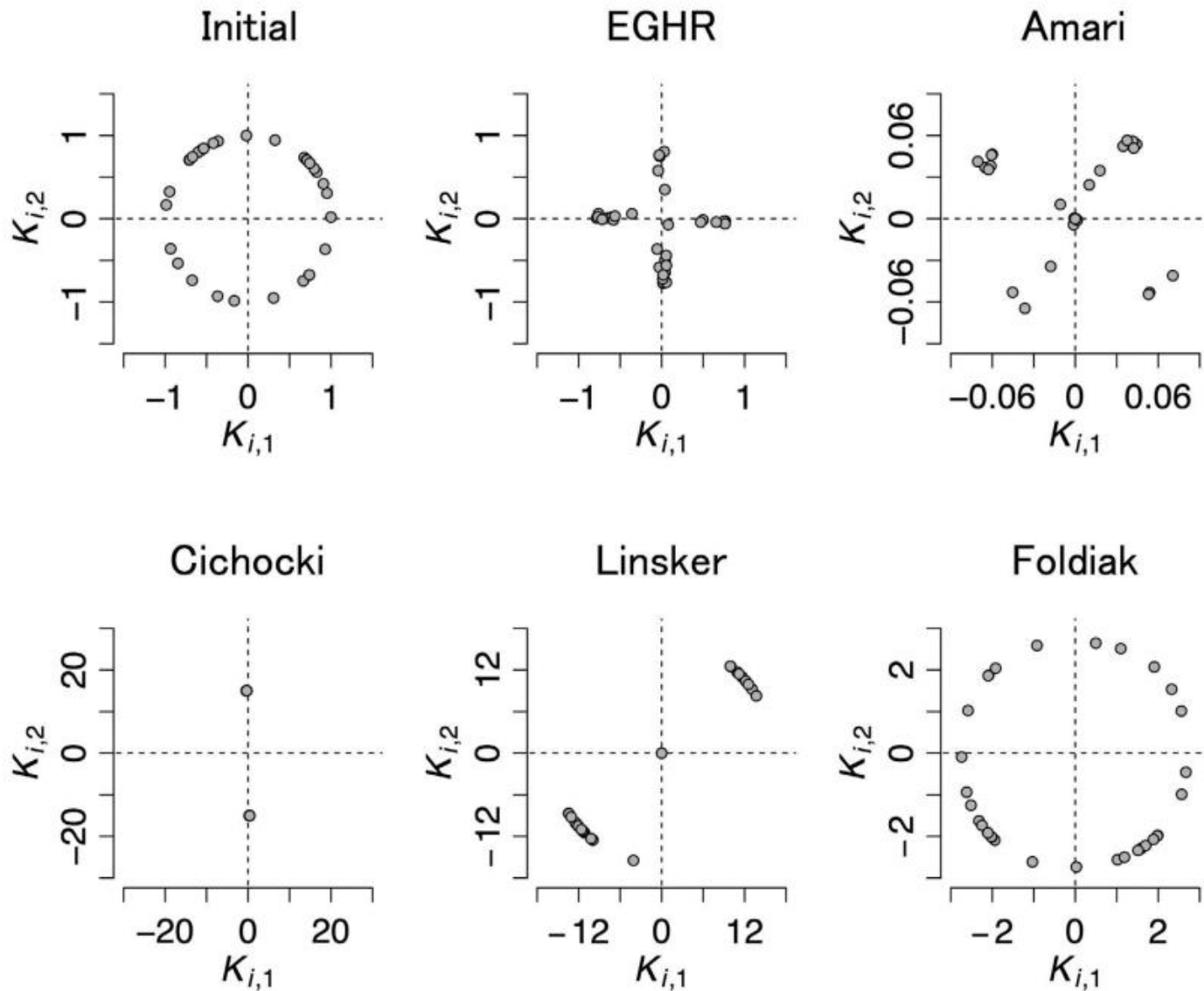
$$E(\mathbf{u}) = -\sum_i \log p_0(u_i)$$



原信号の混ぜ合わせ $\mathbf{x} = \mathbf{A}\mathbf{s}$ と分離 \mathbf{u}

\mathbf{s} はラプラスまたは2重指数分布

$$\frac{1}{2\phi} \exp\left(-\frac{|x - \mu|}{\phi}\right) : \text{期待値 } \mu, \text{ 分散 } 2\phi^2$$



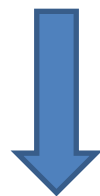
単位円上の点(EGHR)が圧倒的に精度が高い

講義予定の変更

第4回の知能化分科会	6月 9日(土)10:00-12:00
第5回の知能化分科会	6月16日(土)10:00-12:00
第6回の知能化分科会	6月23日(土)10:00-12:00
第7回の知能化分科会	7月 7日(土)10:00-12:00
第8回の知能化分科会	7月14日(土)10:00-12:00
第9回の知能化分科会	7月21日(土)10:00-12:00

「ニューラルネットワークと深層学習第1~4章」

「関数補間」



ただし、「ニューラル
...」に出てくる数字パ
ターンデータ、計算機
コードは参考にする

瀧 雅人「**これならわかる深層学習**」講談社

本文339ページ、価格 ¥3,240 MLS機械学習スタートアップ
シリーズ 2017年10月第1刷発行, 2018年2月第4刷発行

第1章 はじめに

第2章 **機械学習と深層学習**

第3章 **ニューラルネット**

第4章 勾配効果法による学習

第5章 **深層学習の正規化**

第6章 誤差逆伝搬法

第7章 **自己符号化器**

第8章 **畳み込みニューラルネット**

第9章 再帰型ニューラルネット

第10章 ボルツマンマシン

第11章 深層強化学習

付録A 確率の基礎

付録B 変分法

**赤字の部分を
講義する予定**

著者「瀧 雅人博士」

2009年 東京大学大学院理学系研究科
物理学専攻博士後期過程修了

現在 理化学研究所数理創造プログラム(iTHEMS)
上級研究員

2 機械学習と深層学習

2.1 なぜ深層学習か？

ニューラルネットワーク(NN)の研究自体は長い歴史を持っており、起源は1940年代までさかのぼれる。

幾度も過大な期待を寄せられては失望に終わってきたが、**今回のものは一線を画している**。

計算機の実力の飛躍的な向上を背景にして、本格的なNNを計算機に実装できるようになった。

計算機実験をしたら、予想に反して極めて高い性能が得られ、**「NNはうまく行かない」**という長年の思い込みを覆した。

深層学習が何故注目されるのか？それは、計算機に明示的に指示するプログラムの必要がないため。

NNは、人間と同じように学習により問題解決法を学ぶからです。

ブレークスルーにより、深層学習は高い汎化能力の獲得に成功した。

2.1 機械学習とは何か

機械学習とは、人間がこなす学習や知的作業を計算機にさせるための研究および手法を指す。

T. M. ミッチェルの定義：機械学習の基本的な要素は経験 E (データ), タスク T (画像認識), 評価尺度 P (汎化能力) よりなる。学習とは、 E を通して T に対する P の向上。

2.2 代表的なタスク

(1) クラス分類

データをいくつかのカテゴリに仕分ける作業

$$x \rightarrow y(x) \in \{0, 1\}$$

(2) 回帰(関数あてはめ)

過去数日の気象データ x から明日の気温 y (連続値) を予測する作業

$$x \rightarrow y(x) \in \mathbb{R}$$

2.3 さまざまなデータセット

(1) MINIST

MINISTデータベース (Mixed National Institute of Standards and Technology database) はアメリカの国立標準技術研究所 (NIST) が提供する手書き数字のデータベースをシャッフルして作られたデータの集合。手書き数字の画像データが訓練用に6万枚、テスト用に1万枚用意されている。グレースケールの 28×28 ピクセル画像。

(2) ImageNet

写真に写っている特定の物体の分類をさせる物体カテゴリ認識や、一般の画像に写っている物体を検出／分類させる物体検出のための機械学習モデルを開発するための「実験台」として、さまざまな画像データが用意されてきた。ImageNetはその代表的なもので、約1400万枚の画像からなる巨大データベースです。しかも、その画像のそれぞれに、写っている物体一つのカテゴリの正解ラベルがついている。

もう少しこじんまりとした実験を行うときに重宝されているものに、CIFAR-10がある。10のカテゴリに分けられた 32×32 ピクセルの画像データセットである。

2.3 統計入門

2.3.1 標本と検定

統計解析に用いるデータは、母集団から抽出されたものとみなす。データの分析から、データの背後にある母集団についての知識を獲得することを目的とする。疫学者が飲酒量と健康の関係を調べたいとき、地球上すべての人類(母集団)を調査することはできないので、ランダムに抽出した少人数に対する調査結果(データ)を分析することで、人類全てに通用する飲酒量と健康の相関関係を読み取ろうとする。

母集団の性質は、データ生成分布 $P_{data}(x)$ により特徴付けられているものとする。すなわち、確率論的なプロセスにしたがってランダムに生じているとする。

x という具体的なサンプルは、データ成分から抽出されたものとし

$$x \sim P_{data}(x)$$

と仮定する. この式の意味はある変数 x の実現値 x が分布 $P_{data}(x)$ から生じていることを表す.

サンプルは無作為抽出されたもの, より正確には, すべてのサンプルは同一分布から独立に (i.i.d: independently and identically) に抽出されたものである.

母集団の知識を得ることは生成分布を知ることの意味し, データを特徴付けるのに十分な統計量をパラメータ(母数)と呼ぶ. ガウス分布ならば平均値と分散である.

実現象は極めて複雑であり、無数のパラメータを有する。そこで、実分布をよく近似するモデル分布 $P_{data}(x; \theta)$ を仮定し、そのモデルのパラメータ θ の最適値 θ^* をデータから推定する。データ生成プロセスについて推論ができれば、その結果を使うことで、新規のデータに関する予測が可能となる。

2.3.2 点推定

未知の確率分布のパラメータを推定するのが統計的推定という。手持ちのデータ集合 $D = \{x_1, x_2, \dots, x_N\}$ からパラメータの尤もらしい値を推定するのが点推定という。

すなわち，確率変数の関数として推定量

$$\hat{\theta}(x_1, x_2, \dots, x_N)$$

を作る.

推定量に推奨される性質

1. バイアスが小さい

バイアスの定義: 期待値と真値の差

$$b(\hat{\theta}) = E[\hat{\theta}] - \theta^*$$

バイアスがゼロのものを不偏推定量

バイアスがあるものの，データ数が増えるとともにゼロに漸近するものを漸近不偏推定量という.

2. 分散が小さい

真値に対する推定値のばらつき具合

$$\text{Var}(\hat{\theta}) = E[(\hat{\theta} - \theta^*)^2]$$

3. 一貫性

データ点の数が増えるにつれて統計量が真のパラメータに近づく性質.

$$\hat{\theta} \rightarrow \theta^*$$

この性質を満たす推定量を一致推定量という.

(1) ガウス分布

定義 $P(x) = N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

期待値 $E_N[x_n] = \int_{-\infty}^{\infty} x_n P(x_n) dx_n = \mu$

推定量 $\hat{\mu} = \frac{1}{N} \sum_{n=1}^N x_n$

$\hat{\mu}$ は不偏推定量である. 何故なら

$$E_N[\hat{\mu}] = \frac{1}{N} \sum_{n=1}^N E_N[x_n] = \mu$$

分散については $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \hat{\mu})^2$

期待値を計算すると

$$\begin{aligned} \mathbf{E}_N \left[\frac{1}{N} \sum_{n=1}^N (x_n - \hat{\mu})^2 \right] \\ &= \mathbf{E}_N \left[\frac{1}{N} \sum_{n=1}^N \left((x_n - \mu)^2 - 2(x_n - \mu)(\hat{\mu} - \mu) + (\hat{\mu} - \mu)^2 \right) \right] \\ &= \sigma^2 - 2 \frac{\sigma^2}{N} + \frac{\sigma^2}{N} = \left(1 - \frac{1}{N} \right) \sigma^2 \end{aligned}$$

したがって、分散は漸近不偏推定量である。

$$\hat{\hat{\sigma}}^2 = \frac{N}{N-1} \hat{\sigma}^2 = \frac{1}{N-1} \sum_{n=1}^N (x_n - \hat{\mu})^2 \quad \text{が不偏推定量}$$

(2) ベルヌーイ分布

コインの裏表のような2値状態を記述する確率分布である。確率変数 x は0か1かの離散値を取るものとし、 $x=1$ の確率を p とすると、この分布は

$$P(x) = p^x (1-p)^{1-x}$$

となる。ベルヌーイ分布の期待値と分散は

$$E_p[x] = \sum_{x=0,1} xP(x) = 0 \cdot P(0) + 1 \cdot P(1) = p$$

$$E_p[(x-p)^2] = \sum_{x=0,1} (x-p)^2 P(x)$$

$$= (0-p)^2 (1-p) + (1-p)^2 p = p(1-p)$$

パラメータの推定量は $\hat{p} = \frac{1}{N} \sum_{n=1}^N x_n$ とすると

$$E_P[\hat{p}] = \frac{1}{N} \sum_{n=1}^N E_P[x_n] = p \quad \text{だから不偏推定量である.}$$

\hat{p} の分散については

$$E_P[(\hat{p} - p)^2] = \frac{1}{N} \sum_{n=1}^N E_P[(x_n - p)^2] = \frac{1}{N} p(1 - p)$$

であり, データが大きくなると推定値 \hat{p} のばらつきが小さくなる.

2.3.3 最尤推定

データ生成のパラメトリックモデル $P_{\text{model}}(\mathbf{x}; \boldsymbol{\theta})$ が与えられているとし、与えられたサンプル

$D = \{x_1, x_2, \dots, x_N; \boldsymbol{\theta}\}$ はこの分布から無作為に抽出されていると仮定する。これらは同じモデル分布から生成されたものだから、このデータ集合の同時確率密度は

$$P(x_1, x_2, \dots, x_N; \boldsymbol{\theta}) = \prod_{n=1}^N P_{\text{model}}(x_n; \boldsymbol{\theta})$$

となる。これを $L(\boldsymbol{\theta}) = P(x_1, x_2, \dots, x_N; \boldsymbol{\theta})$ と書き、尤度関数と呼ぶ。

データ値が $\{x_1, x_2, \dots, x_N; \boldsymbol{\theta}\}$ を取ったのは、 $L(\boldsymbol{\theta})$ を最大にするパラメータ値であったと考える。

すなわち、データから見て尤もらしいパラメータの値は尤度を最大にするものだと考えられ、パラメータの推定に使いよう。

$$\boldsymbol{\theta}_{ML} = \max L(\boldsymbol{\theta})$$

例) ガウス分布をする N 個のデータの対数尤度関数は

$$\begin{aligned} \log L(\boldsymbol{\theta}) &= \log L(\mu, \sigma^2) = \log \left(\prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n - \mu)^2}{2\sigma^2}} \right) \\ &= -N \log \sigma^2 - \sum_n (x_n - \mu)^2 / 2 + \text{const} \end{aligned}$$

この関数の最大値は、下記の方程式から決まる。

$$0 = \left. \frac{\partial \log L(\mu, \sigma^2)}{\partial \mu} \right|_{ML} \quad 0 = \left. \frac{\partial \log L(\mu, \sigma^2)}{\partial \sigma^2} \right|_{ML}$$

2.4 機械学習の基礎



次週

講義担当者のニューロとの関わり

1990年代に、日立造船技研で、「ニューロによる自動車の車番認識の開発」を行っていた。汎化能力が向上しないので大変苦労した。

当時、ニューロやファジィの勉強をかなりした。
ニューロに関する勉強をまとめて、ニューロの解説書を出版した。

計算機アルゴリズム研究会、「C言語による実用ニューロ・コンピューティング」、(株)ラッセル社、
(1992)、¥2,800, 223ページ

目次

はじめに

1. 情報処理の2つの方式

- 1.1 現在の計算機による情報処理
ノイマン型計算機の特徴と限界
- 1.2 人間の脳における情報処理
ニューロ・コンピューティング
..... 高度並列分散処理システム

2. さまざまなニューロ・コンピューティング

- 2.1 簡単なネットワーク(神経回路)
- 2.2 神経細胞の数理モデル
- 2.3 神経回路の学習規則

2.4 相互結合型ニューラル・ネットワーク

2.4.1 アソシアトロンとその検索システムへの応用

2.4.2 ホップフィールド・マシン

2.4.3 ボルツマン・マシン

2.5 階層構造型ニューラル・ネットワーク

2.5.1 パーセプトロン

2.5.2 誤差逆伝播学習型多層パーセプトロン

3. 誤差逆伝播学習型多層パーセプトロン

3.1 学習理論

3.2 アルゴリズムについて

3.3 簡単な誤差逆伝播学習型パーセプトロン

3.4 簡単な誤差逆伝播学習型パーセプトロンの応用

4. 誤差逆伝播学習型多層パーセプトロンの実際

4.1 C言語の簡単な説明

4.2 プログラムの説明

4.3 コンパイル, リンク, 実行

4.4 実行結果

4.5 チューニング

5. 学習済みネットワークによるパターン認識

5.1 プログラムの説明

5.2 コンパイル, リンク, 実行

5.3 実行結果

5.4 機能の拡張について

終わりに

参考文献

6. ペーパーウェア

AETRON... 検索システムのプログラム(アソシアトロン)

PTRN ... 2層ネットワーク(パーセプトロン)

TINY ... 簡単な誤差逆伝播学習型3層パーセプトロン

STUDY ... 誤差逆伝播学習型4層パーセプトロンによる
数字パターン学習

JUDGE ... 学習済のネットワーク(4層パーセプトロン)
による数字パターン認識

内容の一部

2.1 簡単なネットワーク(神経回路)

出力層	出力	0	1	1	1
	入力	0	1	1	2
入力層	出力	0 0	1 0	0 1	1 1

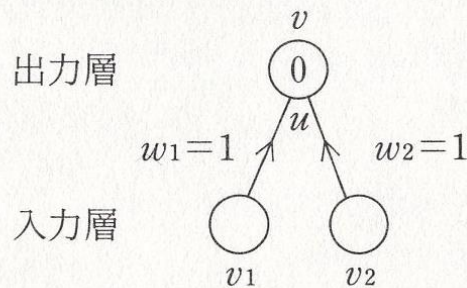


図2.1.1 論理和(OR)を実現するネットワーク

出力層	出力	0	0	0	1
	入力	0	1	1	2
入力層	出力	0 0	1 0	0 1	1 1

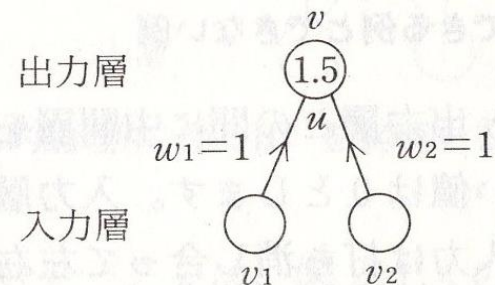


図2.1.2 論理積(AND)を実現するネットワーク

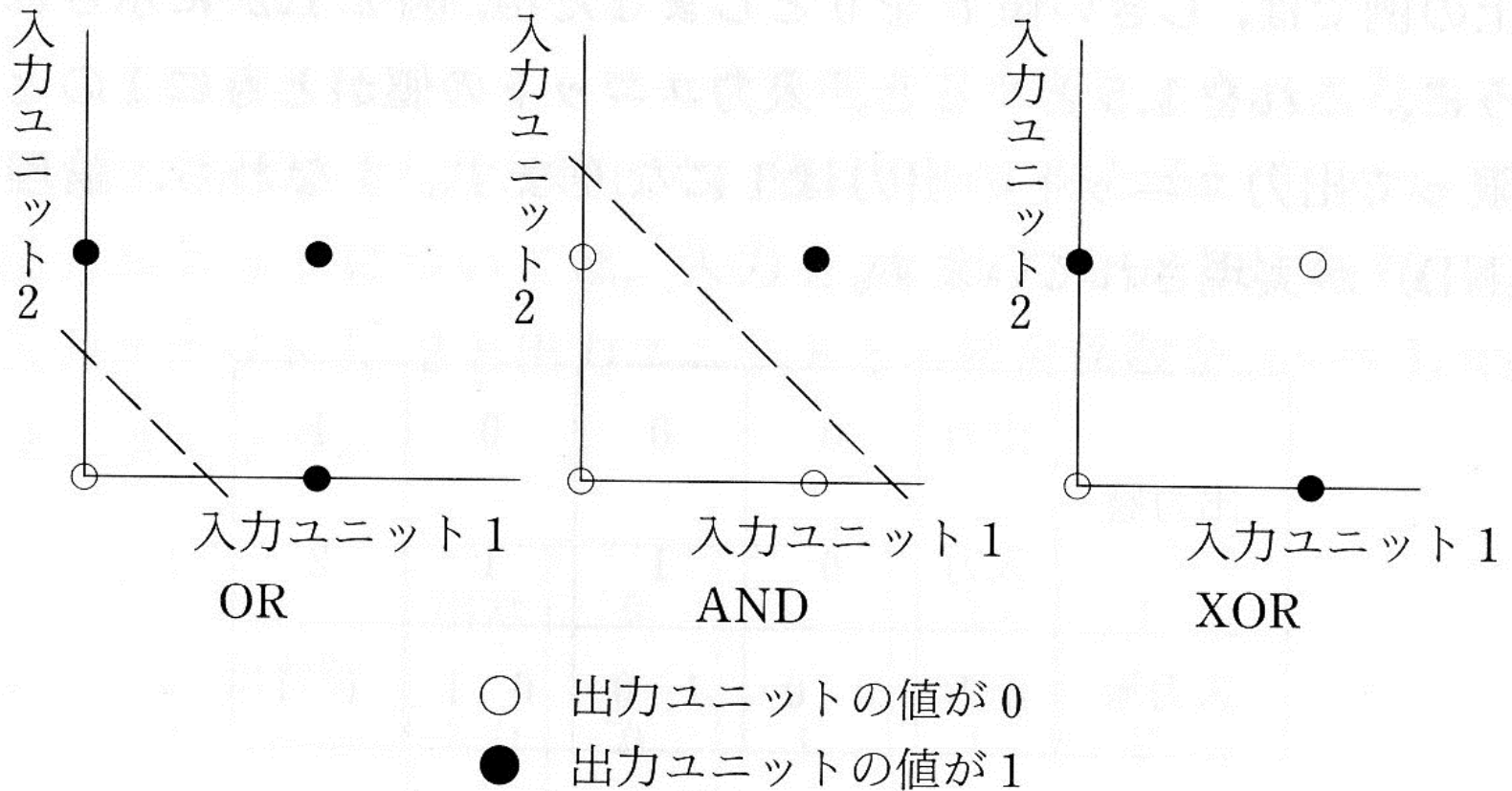


図2.1.3 線型分離できる例とできない例

出力層	出力	0	1	1	0
	入力	0	1	1	0
中間層	出力	0 0	1 0	0 1	0 0
	入力	0 0	1 -1	-1 1	0 0
入力層	出力	0 0	1 0	0 1	1 1

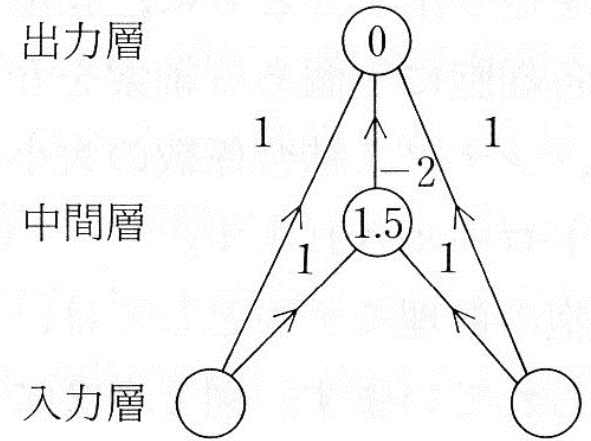


図2.1.5 排他的論理和 (XOR)を実現するネットワーク2

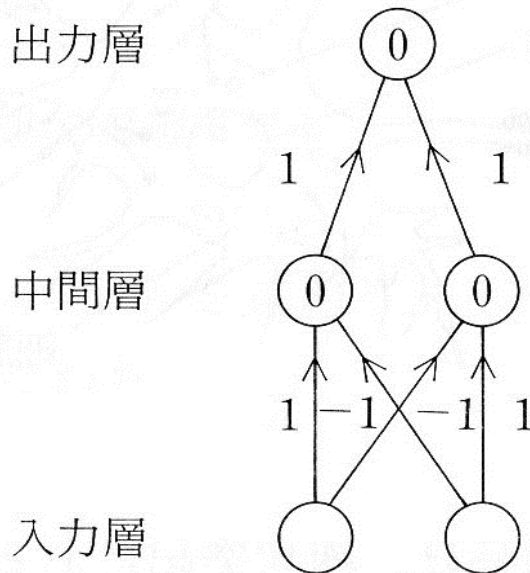
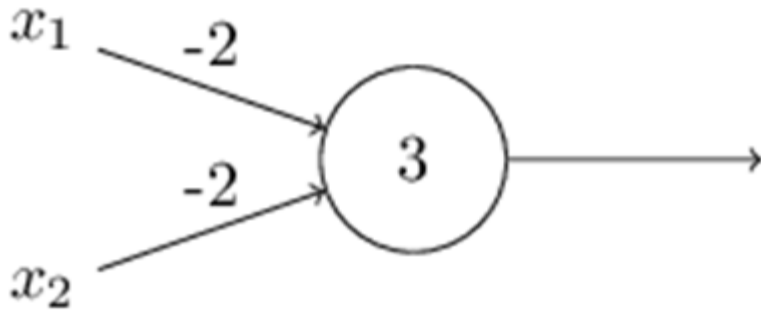


図2.1.4 排他的論理和 (XOR)を実現するネットワーク1

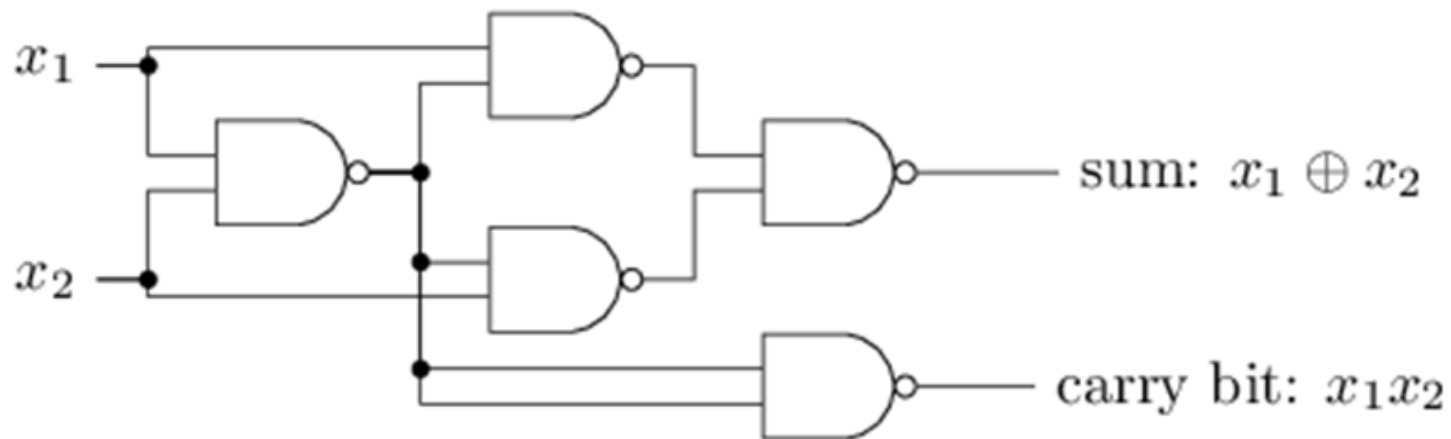
追加



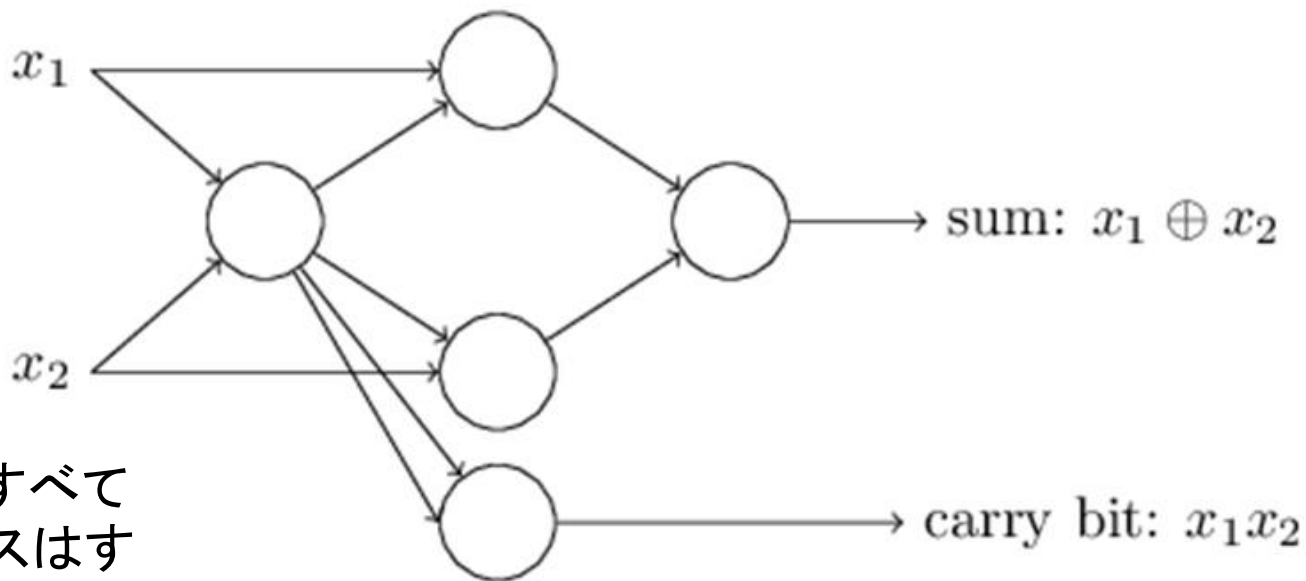
否定論理積 (NAND) の
ネットワーク

入 力		出 力
x_1	x_2	y
0	0	1
1	0	1
0	1	1
1	1	0

NANDを用いればあらゆる論理回路が可能になる。
→ ニューラルネットワークであらゆる論理回路が可能
(NAND論理の完全性)



NANDゲートによる2進数の和の実装(電子回路)



ウェイトはすべて
-2、バイアスはす
べて3

ニューロによる2進数の和の実装

3.4 簡単な誤差逆伝播の応用

[例題1] 試合結果の判定

Aさん, Bさん, Cさん, Dさんの4人がそれぞれダブルスを組み, 残りの2人(ダブルス)を相手にそれぞれ10回ずつテニスの試合を行いました.

表3.4.1 ダブルスの組み合わせと対戦試合結果

組合せ	A	B	C	D	勝	負
①	○	○			9	1
②	○		○		4	6
③	○			○	7	3
④		○	○		4	6
⑤		○		○	4	6
⑥			○	○	2	8

表3.4.2 各人の試合の組み合わせと対戦結果

	試合の組合せ			勝	負
A	①	②	③	20	10
B	①	④	⑤	17	13
C	②	④	⑥	10	20
D	③	⑤	⑥	13	17

表3.4.3 ダブルスの組み合わせの大戦試合
結果の書き直し

組合せ	A	B	C	D	勝	負
①	1.0	1.0	0.0	0.0	0.9	0.1
②	1.0	0.0	1.0	0.0	0.4	0.6
③	1.0	0.0	0.0	1.0	0.7	0.3
④	0.0	1.0	1.0	0.0	0.4	0.6
⑤	0.0	1.0	0.0	1.0	0.4	0.6
⑥	0.0	0.0	1.0	1.0	0.2	0.8

各人の頑張り度計算のための入力データ

入力値 ユニット 0 A	入力値 ユニット 1 B	入力値 ユニット 2 C	入力値 ユニット 3 D
1.0000	0.0000	0.0000	0.0000
0.0000	1.0000	0.0000	0.0000
0.0000	0.0000	1.0000	0.0000
0.0000	0.0000	0.0000	1.0000

各人の頑張り度

出力値 ユニット 0 勝	出力値 ユニット 1 負
0.8974	0.1028
0.6889	0.3047
0.2511	0.7418
0.3071	0.6898

Aさん, Bさん, Dさん, Cさんの順で頑張っている.

飲み会の代わり...自由討論

分科会運営方法に関するコメント

最近の世相，時代の潮流

夢：遠隔地音楽アンサンブル

時間遅れをどう克服するか？

音楽合奏に関する鍋島さんの経験

ユニゾンからアンサンブルへ