

# 導入

杉山 将

sugi@k.u-tokyo.ac.jp

<http://www.ms.k.u-tokyo.ac.jp>

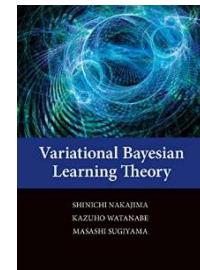
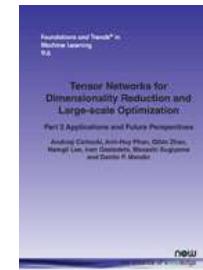
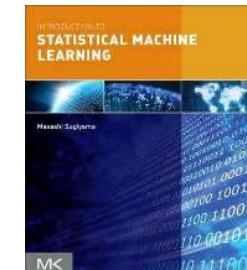
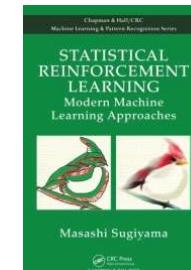
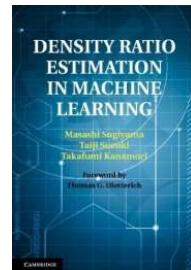
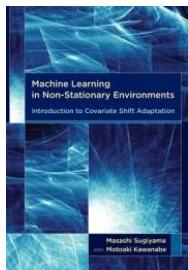
# 自己紹介

## ■ 現職：

- 理化学研究所・センター長：研究者とともに
- 東京大学・教授：学生とともに
- 企業・技術顧問：エンジニアとともに

## ■ 専門分野：

- 人工知能の数学的な基礎研究  
(コンピュータ, 統計学など)
- 人工知能技術の実世界応用  
(画像, 言語, 脳波, ロボット, 医療, 生命など)



# 機械学習

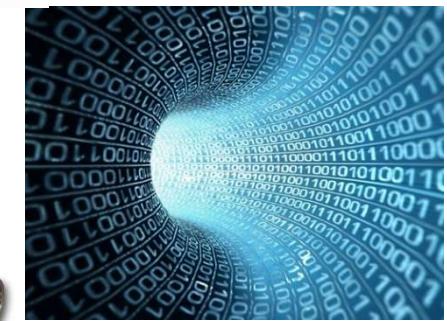
■ 機械学習：データの背後に潜む知識を学習する

■ 様々な応用例：

- 音声・画像・動画の認識
- ウェブやSNSからの情報抽出
- 商品やサービスの推薦
- 工業製品の品質管理
- ロボットシステムの制御



■ ビッグデータ時代の到来に伴い、  
機械学習技術の重要性は  
益々高まりつつある



# 先端データ解析論

- **内容**: 機械学習の主要技術を紹介する
- **担当**: 杉山将, 本多淳也
- **成績評価**:
  - ほぼ毎週出す小レポート課題  
(次の講義までにITC-LMSで提出)
  - 講義中にたまに行う小さな演習  
(発表した学生には加点する)
- **講義資料**:
  - ITC-LMSのサイトで公開する

# 講義スケジュール

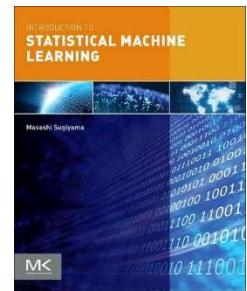
1. 4/9: 概要
2. 4/16: 最小二乗回帰
3. 4/23: スパース回帰
4. 5/7: ロバスト回帰
5. 5/14: 最小二乗分類
6. 5/21: サポートベクトル分類
7. 5/28: ロジスティック回帰
8. 6/4: オンライン分類
9. 6/11: 休講
10. 6/18: 半教師付き学習  
・転移学習
11. 6/25: 教師なし線形  
次元削減
12. 7/2: 教師付き線形  
次元削減
13. 7/9: カーネル非線形  
次元削減
14. 7/16: 発展的な話題

6/4は情報理工講義日でないが、  
新領域は講義日なので講義を行う

# 教科書

## ■ 杉山: イラストで学ぶ機械学習, 講談社, 2013

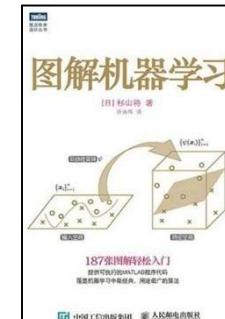
- 最小二乗回帰, スパース回帰, ロバスト回帰
- 最小二乗分類, サポートベクトル分類, 確率的分類, 系列データの分類
- オンライン学習, マルチタスク学習, 半教師付き学習, 転移学習
- 線形次元削減, カーネルを用いた非線形次元削減, ニューラルネットワークを用いた非線形次元削減



## ■ Sugiyama, M.: Introduction to Statistical Machine Learning, Morgan Kaufmann, 2015.

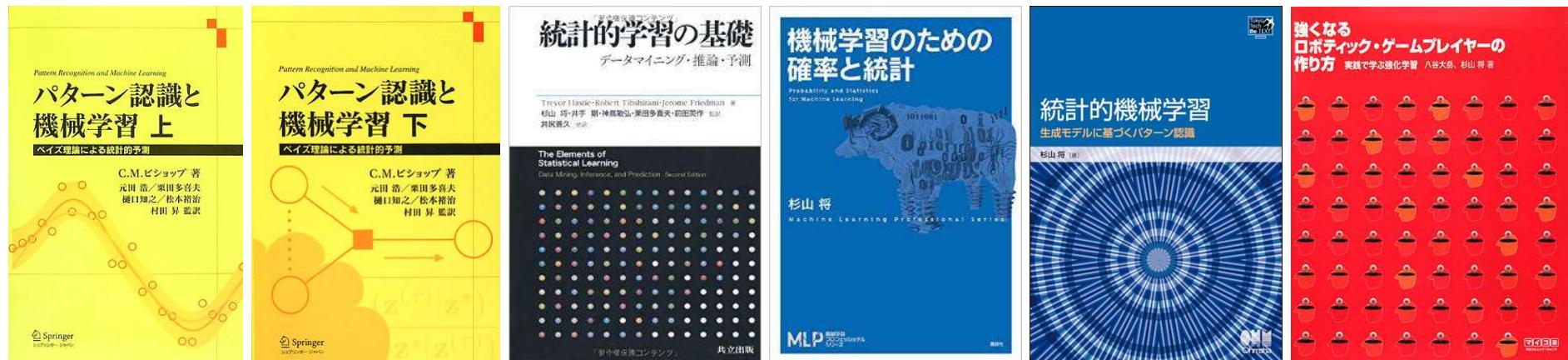
## ■ 杉山: 图解机器学习

## ■ 스기야마: 그림과 수식으로 배우는 통통 머신러닝



# 参考書

- Bishop著, 元田ほか訳:パターン認識と機械学習(上・下),  
丸善出版, 2007.
- Hastieほか著, 杉山ほか訳:統計的学习の基礎, 共立出版, 2014.
- 杉山, 機械学習のための確率と統計, 講談社, 2015.
- 杉山:統計的機械学習, オーム社, 2009.
- 八谷, 杉山:強くなるロボティック・ゲームプレイヤーの作り方,  
毎日コミュニケーションズ, 2008.



# その他日本語の参考書

## 講談社:機械学習プロフェッショナルシリーズ(全30巻)



- 機械学習のための連続最適化:  
金森敬文, 鈴木大慈, 竹内一郎, 佐藤一誠
- オンライン予測:畠埜晃平, 瀧本英二
- 関係データ学習:石黒 勝彦, 林 浩平
- データ解析におけるプライバシー保護:佐久間 淳
- ウェブデータの機械学習:  
ダヌシカ ボレガラ, 岡崎直觀, 前原貴憲
- バンディット問題の理論とアルゴリズム:  
本多淳也, 中村篤祥
- グラフィカルモデル:渡辺有祐
- ヒューマンコンピュテーションとクラウドソーシング:  
鹿島久嗣, 小山聰, 馬場雪乃
- ノンパラメトリックベイズ 点過程と統計的機械  
学習の数理:佐藤一誠
- 変分ベイズ学習:中島伸一
- スパース性に基づく機械学習:富岡亮太
- 生命情報処理における機械学習多重検定と  
推定量設計:瀬々潤, 浜田道昭
- 劣モジュラ最適化と機械学習:河原吉伸, 永野清仁
- 統計的学習理論:金森敬文
- 確率的最適化:鈴木大慈
- 異常検知と変化検知:井手剛, 杉山将
- サポートベクトルマシン:竹内一郎, 烏山昌幸
- 機械学習のための確率と統計:杉山将
- 深層学習:岡谷貴之
- オンライン機械学習:  
海野裕也, 岡野原大輔, 得居誠也, 徳永拓之
- トピックモデル:岩田具治
- 統計的因果探索:清水昌平
- 画像認識:原田達也
- 深層学習による自然言語処理:坪井祐太,  
海野裕也, 鈴木潤
- 音声認識:篠田浩一
- ガウス過程と機械学習:持橋大地, 大羽成征
- 強化学習:森村哲郎
- ベイズ深層学習:須山敦志
- ロボットの運動学習(刊行予定)
- 統計的音響信号処理(刊行予定)



# 講義の流れ

## 1. 研究業界の動向

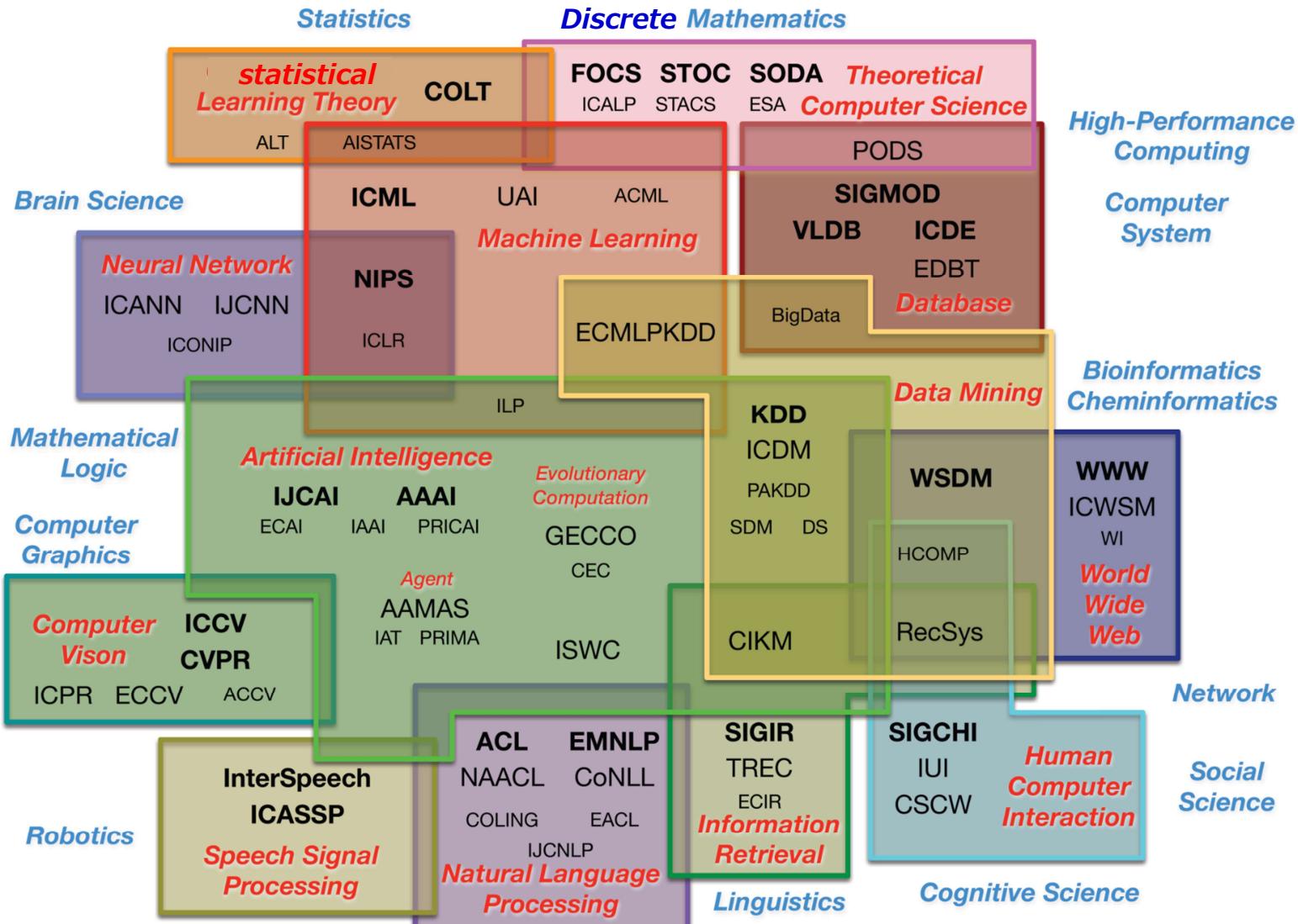
- A) 国際会議
- B) 日本
- C) 理研

## 2. 機械学習の基礎

## 3. 確率・統計の基礎

# 「人工知能」に関する国際会議<sup>10</sup>

■ 基礎数学から実世界応用まで、様々な学会が存在



※産総研・神鳶先生資料より抜粋（一部改変）

ML, DM, and AI Conference Map, Copyright © 2015 Toshihiro Kamishima All Rights Reserved, Updated 2015/02/22

# 機械学習の国際会議の動向

11

## ■ 参加者数が激増：

	2013	2014	2015	2016	2017	2018
ICML	900	1200	1600	3000+	2400	5000
NIPS	1200	2400	3800	6000+	7500+	8000+

- ICML: International Conference on Machine Learning
- NIPS (NeurIPS): Neural Information Processing Systems

## ■ 企業のスポンサーも非常に活発：

- 2000年台前半～：アメリカのIT企業 (Google, IBM, Yahoo!, Microsoft...)
- 2000年台後半～：世界中のIT企業 (Amazon, Facebook, Linkedin, Tencent, Baidu, Huawei, Yandex...)
- 2010年台：製造・金融など様々な業種のスタートアップ～大企業

# 研究業界の動向

## ■技術開発: アメリカ一強

- カナダ, ドイツ, イギリス, フランスなどにも  
アメリカ系IT企業が次々と研究所を開設

## ■ビジネス: アメリカと中国が支配的

- アメリカと同種の企業が中国にもあり,  
より大きなマーケットを専有しつつある

## ■日本の存在感は薄い:

- 日本の国際会議での論文占有率は数%程度
- 日本人の研究者数も数%程度
- 中国・韓国は、欧米の大学・企業に所属している  
学生・研究者が多数いるため,  
コミュニティに深く溶け込んでいる



# 講義の流れ

13

## 1. 研究業界の動向

- A) 國際會議
- B) 日本
- C) 理研

## 2. 機械学習の基礎

## 3. 確率・統計の基礎

# 人工知能技術戦略会議

平成28年4月開催の「未来投資に向けた官民対話」での総理指示

**人工知能の研究開発目標と産業化のロードマップを、本年度中に策定します。**

そのため、产学研官の叡智を集め、縦割りを排した『人工知能技術戦略会議』を創設します。



## 人工知能技術戦略会議の設置（平成28年4月）

総理指示を受け、『人工知能技術戦略会議』が創設。同会議が司令塔となって、総務省、文部科学省、経済産業省が所管する5つの国立研究開発法人を束ね、AI技術の研究開発を進めるとともに、AIを利用する側の産業（いわゆる出口産業）の関係府省と連携し、AI技術の社会実装を推進。

議長：安西祐一郎 日本学術振興会理事長

顧問：久間和生 内閣府総合科学技術・イノベーション会議常勤議員

構成員：経団連未来産業・技術委員長、東京大学総長、大阪大学総長、  
ICT理事長、理研理事長、産総研理事長、JST理事長、NEDO理事長

# 人工知能技術戦略(平成29年3月策定)

## 「人工知能技術戦略」(平成29年3月策定) 概要

1. A I 開発関係官庁（総務、文科、経産）が連携し、我が国が有する現場の強みを踏まえ、研究開発から社会実装まで一貫した取組の加速。
2. 内閣府のS I Pを含め、厚生労働省、国土交通省、農林水産省など出口産業を所管する関係府省のプロジェクトと連携。A I 技術の研究開発について民間投資を促進。
3. 重点分野（「生産性」、「健康、医療・介護」、「空間の移動」）における産業化ロードマップの策定。

## (参考) 産業化ロードマップ概要

- 我が国が世界をリードしていくために、我が国や世界が直面している社会課題に対して、我が国が有する現場の強みをも踏まえ、A I 技術とその他関連技術による産業化に向けたチャレンジングなロードマップを掲げて、产学研官の叡智を結集し、研究開発から社会実装まで一貫した取組を加速していく
- 生産性、健康/医療・介護、空間の移動を重点3分野として設定。また、横断的な分野として情報セキュリティも設定。

## (参考) 研究開発目標概要

- 産業化ロードマップを踏まえ、特に国立研究開発法人として中心となって取り組むべき研究開発テーマについて、3センター（N I C T、理研、産総研）は連携して取り組む。
- 3センターが連携して取り組むテーマは以下の観点から選定する。
  - 基礎研究から社会実装まで一貫して取り組むべきもの。
  - 短期的な収益化が見込めず、民間だけでは開発が進まないもの。
  - 国際標準化、共通基盤技術など協調領域であるもの。

# 人工知能技術戦略会議の司令塔機能の強化

**人工知能技術戦略会議を基礎研究から社会実装まで取り組む司令塔**とし、これまでの事務局（総務省、文部科学省、経済産業省の3省）に内閣府、厚生労働省、農林水産省、国土交通省を追加（平成29年12月）。産業化ロードマップの実現に向けて、SIP/PRISMとも連動し、**府省連携**して取り組む。



# 官民研究開発投資拡大プログラム（PRISM）による府省連携の推進

平成30年度政府予算案 100億円（新規）

## 戦略的イノベーション創造プログラム（SIP）

- 府省・産学官連携
- 基礎研究から事業化・実用化まで一気通貫
- 出口戦略の明確化
- 独自の厳格なマネジメントの導入
- 多くの成果／産業界からの高い評価

H26創設

「一本立ての施策」として  
一  
体  
的  
・  
戦  
略  
的  
な  
運  
用

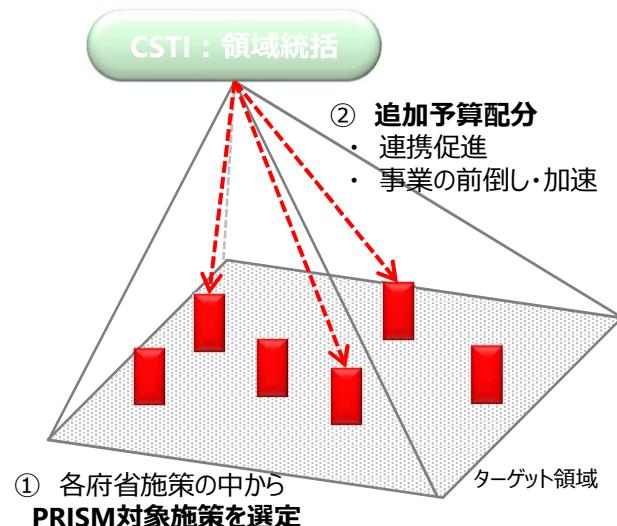
## 官民研究開発投資拡大プログラム（PRISM）

- 民間研究開発投資誘発効果の高い領域（ターゲット領域）を設定
- 総合科学技術・イノベーション会議（CSTI）が追加の予算を配分することにより
  - ✓ 各府省の施策をターゲット領域に誘導
  - ✓ 各府省施策の連携を強力に推進
- SIPの優れた特徴（厳格なマネジメント等）を各府省に展開

H30創設予定

➡ SIP／PRISMによる先導的な取組を、生産性の抜本的向上が必要な分野（農業、建設、ものづくり、物流等）を中心に行き、Society5.0の実現に貢献（AI、センサー、ロボット等のサイバー・フィジカル空間の共通基盤／ビッグデータ統合基盤の構築）

### ◎ PRISMマネジメント体制



### ◎ 平成30年度に創設予定のターゲット領域／領域統括

#### サイバー空間基盤技術

【領域統括】  
**安西 祐一郎**  
日本学術振興会理事長

【主要分野】  
✓ AI/IoT/ビッグデータ

- ・重点3分野：生産性・サービス、健康/医療・介護、空間の移動
- ・研究開発官庁のみならず、事業官庁を含めた体制整備

#### フィジタル空間基盤技術

【領域統括】  
**佐相 秀幸**  
富士通研究所顧問

- 【主要分野】
- ✓ センサー
  - ✓ エッジコンピューティング
  - ✓ アクチュエータ/ロボティクス
  - ✓ 共通基盤（ICT基盤、量子暗号等）

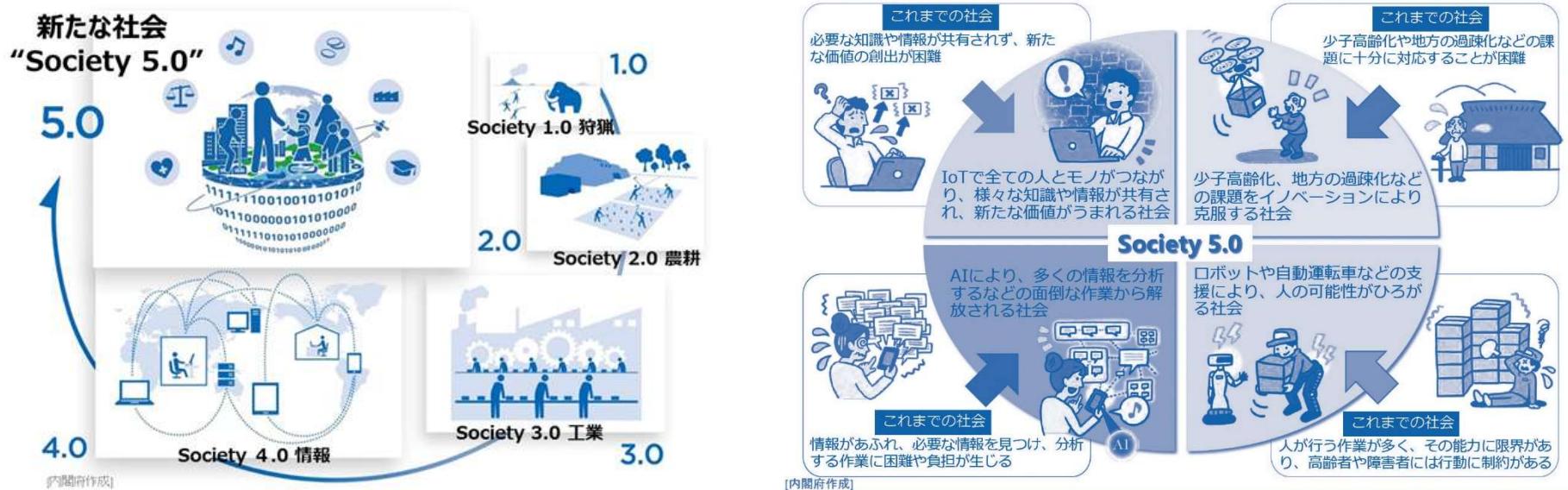
#### 建設・インフラ／防災・減災

【領域統括】  
**田代 民治**  
鹿島建設代表取締役副社長

- 【主要分野】
- ✓ 建設：i-Construction
  - ✓ インフラ維持管理：点検、診断、補修、長寿命化
  - ✓ 防災・減災：予防、予測、被害軽減、早期機能復旧

# Society 5.0: 日本が目指す社会

■ サイバー空間とフィジカル空間を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する**人間中心の社会**



出典:「Society 5.0『科学技術イノベーションが拓く新たな社会』説明資料」(内閣府)



# 講義の流れ

19

## 1. 研究業界の動向

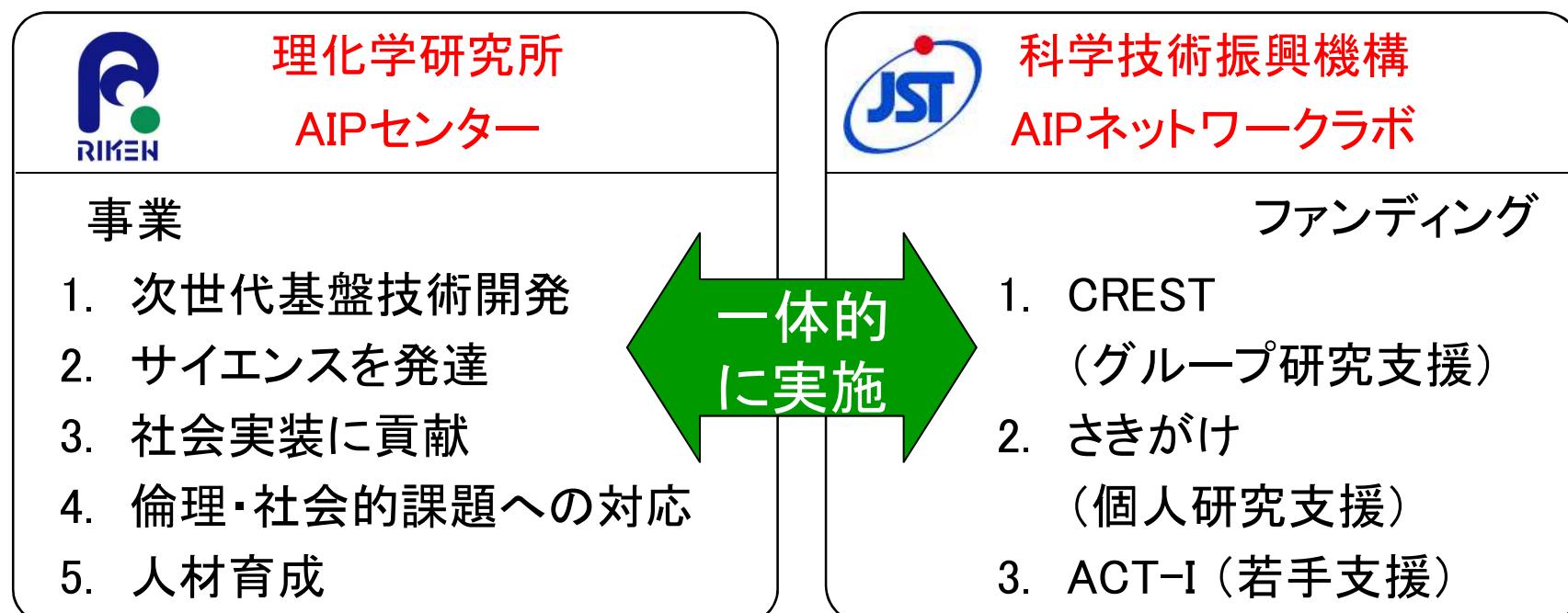
- A) 國際會議
- B) 日本
- C) 理研

## 2. 機械学習の基礎

## 3. 確率・統計の基礎

# 文科省AIPプロジェクト： 人工知能/ビッグデータ/IoT/ サイバーセキュリティ統合プロジェクト

■ 理化学研究所は、2016年4月14日付で  
革新知能統合研究(AIP)センターを設置



# Society 5.0の実現にむけて

- Society 5.0の実現には、  
**人工知能(AI)技術が不可欠！**



AI × 交通、医療・介護、  
      ものづくり、農業、  
      防災、エネルギーなど

様々な分野でAIの活用が始まる

- AIの活用を更に加速していくためには、  
**AI技術そのものも強化していく必要がある**
  - コンピュータにヒトのような学習能力を持たせる  
**機械学習(Machine Learning)**が中心

# AIPセンターの取り組み

ML



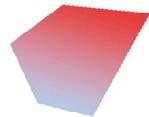
交通, 医療・介護,  
ものづくり, 農業,  
防災, エネルギーなど



深層学習の仕組みの解明や、新しい原理に基づく次世代人工知能技術の創出を目指す  
汎用基盤技術研究グループ



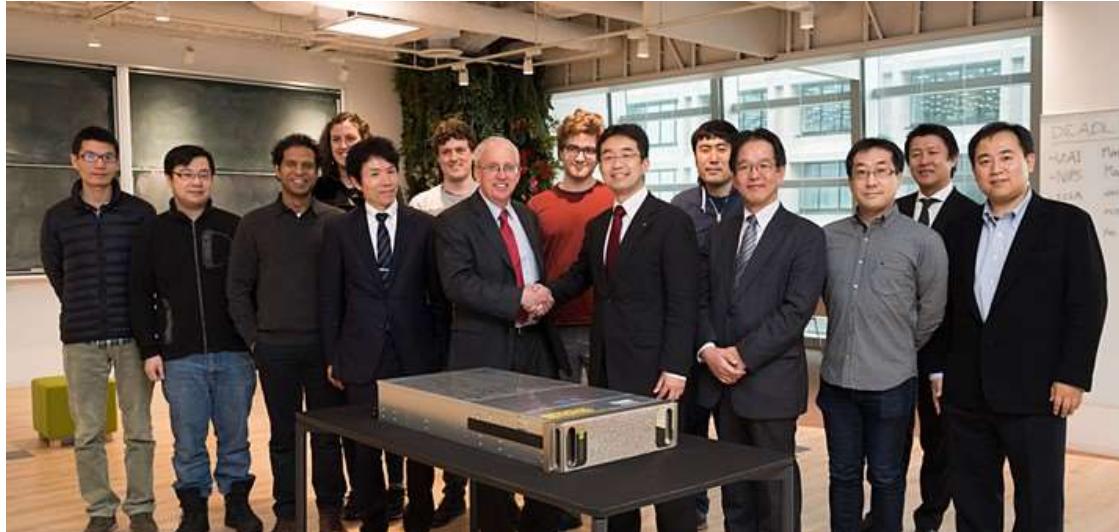
大学・研究機関や産業界と連携し、社会的・経済的価値の創造へ貢献する  
目的志向基盤技術研究グループ



倫理的・法的・社会的影响を分析し、制度に関する提言を行う  
社会における人工知能研究グループ

# 計算リソース: RAIDEN

## (Riken Alp Deep learning Environment)



NVIDIA副社長ビル・ダリー氏と  
<https://blogs.nvidia.co.jp/2017/03/06/fujitsu-ai-supercomputer/>



### ■ NVIDIA DGX-1 Tesla V100

- 半精度54PFLOPS

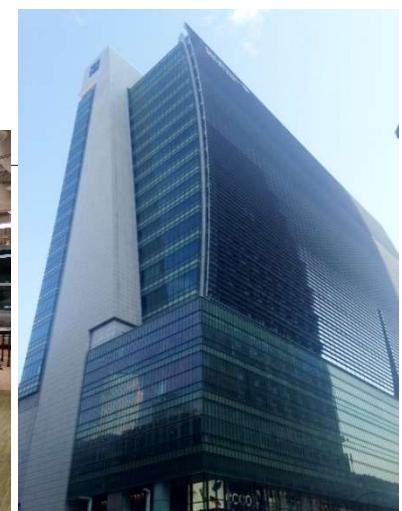
稼働率80%

### ■ 2018年6月のGreen500にて世界10位の電力効率

- 11.363GFLOPS/W

# 全国展開

■ 日本橋を中心に、  
全国各地にチームを配置



# AIPの研究者による東大講義

25

## ■ Special Topics in Mechano-Informatics II (Wed-4)

- |  |  |   |   |
|--|--|---|---|
| 1. 2019/4/10 Masashi Sugiyama:<br>Overview of AI Research<br>and Introduction of<br>RIKEN Center for AIP |    | 7. 2019/5/29 Kentaro Inui:<br>Natural Language<br>Understanding and<br>Commonsense Reasoning                        |    |
| 2. 2019/4/17 Tatsuya Harada:<br>Image and Video Recognition<br>Using Machine Learning                    |    | 8. 2019/6/5 Koji Tsuda:<br>Materials Informatics  |    |
| 3. 2019/4/24 Mihoko Otake:<br>Cognitive Behavioral<br>Assistive Technology                               |    | 9. 2019/6/12 Emtiyaz Khan:<br>Bayesian Deep Learning  |    |
| 4. 2019/5/8 Hiroshi Nakagawa:<br>Social Impact of AI   |  | 10. 2019/6/19 Ichiro Takeuchi:<br>Statistical Inference for<br>Big Data Era: Introduction to<br>Selective Inference |   |
| 5. 2019/5/15 Kohei Hatano:<br>Online Prediction  |  | 11. 2019/6/26 Junya Honda:<br>Algorithms for Bandit Problems  |  |
| 6. 2019/5/22 Kazuki Yoshizoe:<br>Search Algorithms in AI   |  | 12. 2019/7/3 Yoshinobu Kawahara:<br>Learning and Analysis of<br>Dynamical Systems                                   |  |



# 講義の流れ

26

1. 研究業界の動向
2. 機械学習の基礎
3. 確率・統計の基礎



# 3種類の機械学習

27

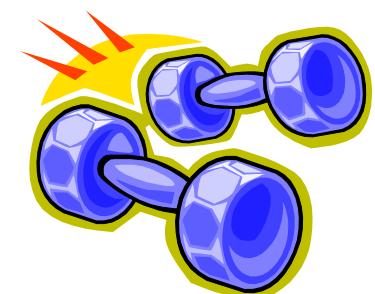
■ 教師付き学習：人間が教師となり、コンピュータの学習を手伝う



■ 教師なし学習：コンピュータが人間の手を介さずに自発的に学習する

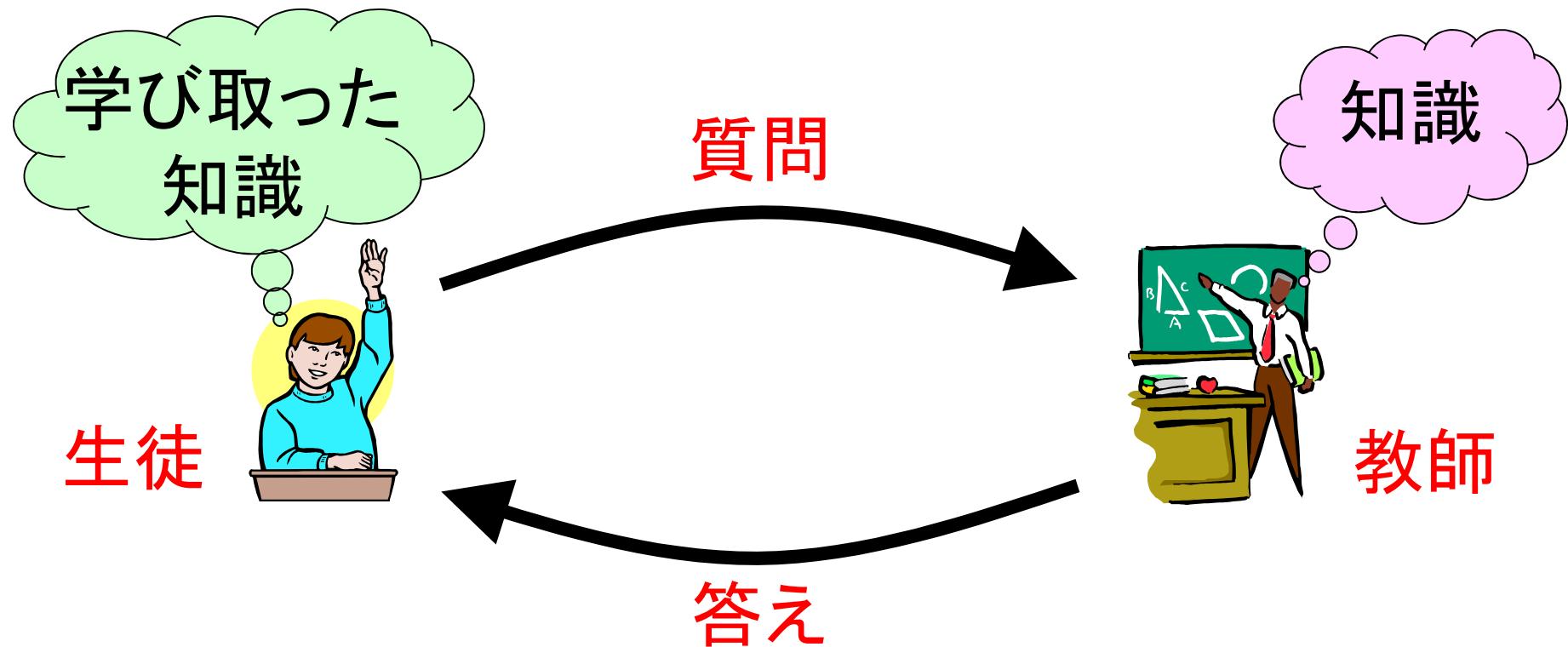


■ 強化学習：ロボットが人間の手を介さずに自発的に意思決定を行う



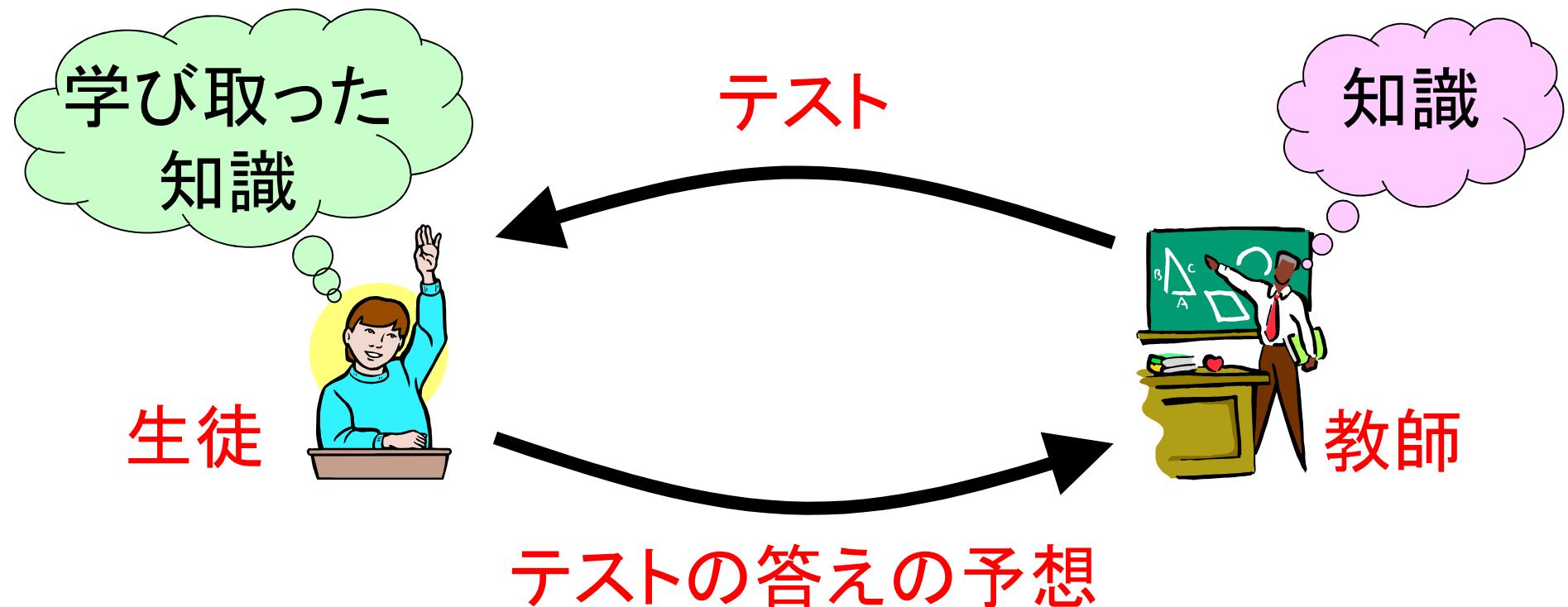
# 教師付き学習

- 教師: どんな質問に対しても答えられる
- 生徒: 教師に質問し、答えを教えてもらう
- ゴール: (生徒が) 教師の知識を学び取る



# 汎化能力

- 教わった答えを覚えるだけでなく、教師の知識**全体**を学び取れば、答えを教わっていない新しい質問（テスト）に対しても正しく解答できる

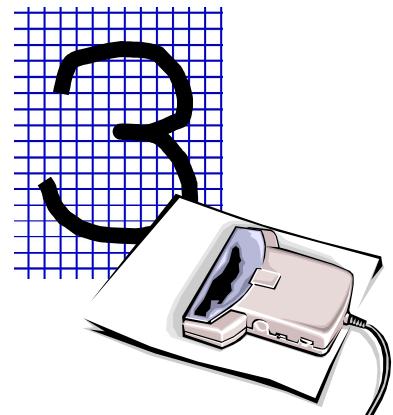


# 例：手書き数字認識

30

## ■ 手書き数字の認識

スキャナで取り込んだ  
文字をコンピュータに  
認識させる

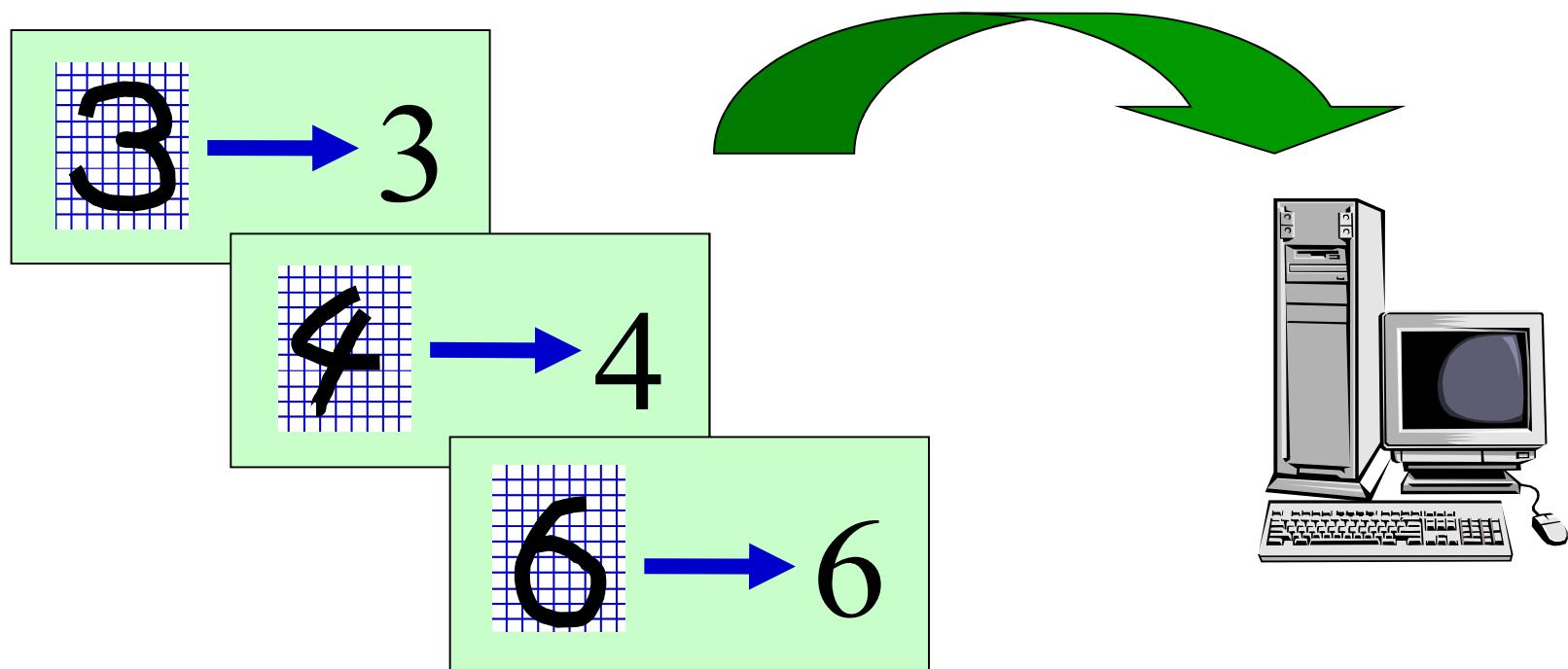


3	6	8	1	7	9	6	6	9	1
6	7	5	7	8	6	3	4	8	5
2	1	7	9	7	1	2	8	4	6
4	8	1	9	0	1	8	8	9	4
7	6	1	8	6	4	1	5	6	0
7	5	9	2	6	5	8	1	9	7
2	2	2	2	3	4	4	8	0	
0	2	3	8	0	7	3	8	5	7
0	1	4	6	4	6	0	2	4	3
7	1	2	8	1	6	9	8	6	1

# 例：手書き数字認識（続き）

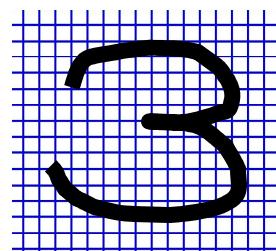
31

- 教師は、数字（の画像）を見たら、それが数字のいくつかを答えることができる。
- 数字の画像の例をいくつかコンピュータに学習させる



# 例：手書き数字認識（続き）

- コンピュータのメモリは大きいので、すべての文字を覚えさせるだけでよいのでは？
- ひとつの文字が $10 \times 10$ 画素で構成され、各画素は256階調の色の濃さを表しているとする。



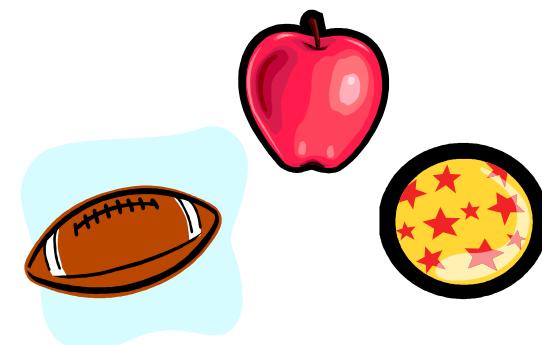
## 例：手書き数字認識（続き）

- すべての文字を覚えさせるためには、  
 $256^{100}$  種類の手書き文字データが必要！

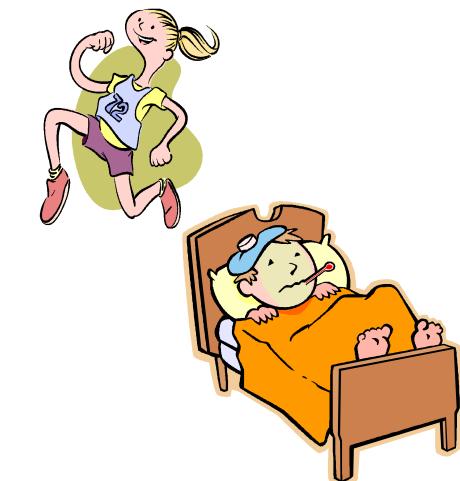
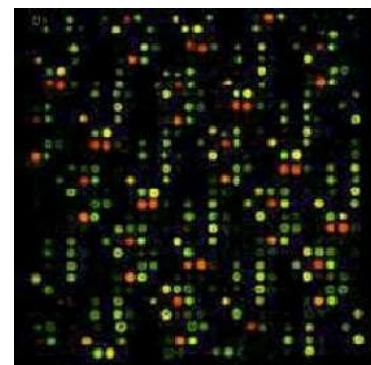
$$256^{100} = (2^8)^{100} = (2^{10})^{80} \approx (10^3)^{80} = 10^{240}$$

- すべてを覚えさせるのは不可能
- すべてのデータを人間が準備するのは  
もっと不可能！

## ■ ロボットビジョン

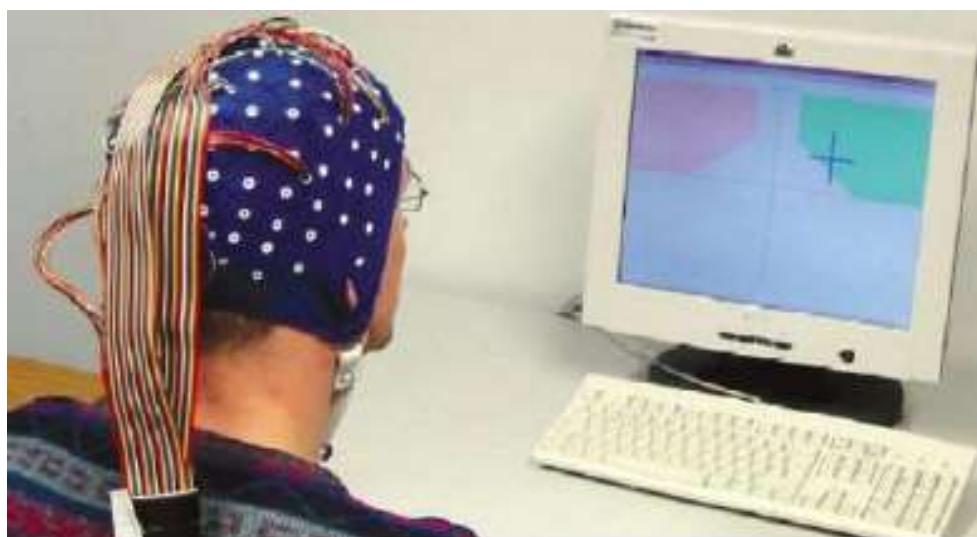


## ■ 遺伝子診断



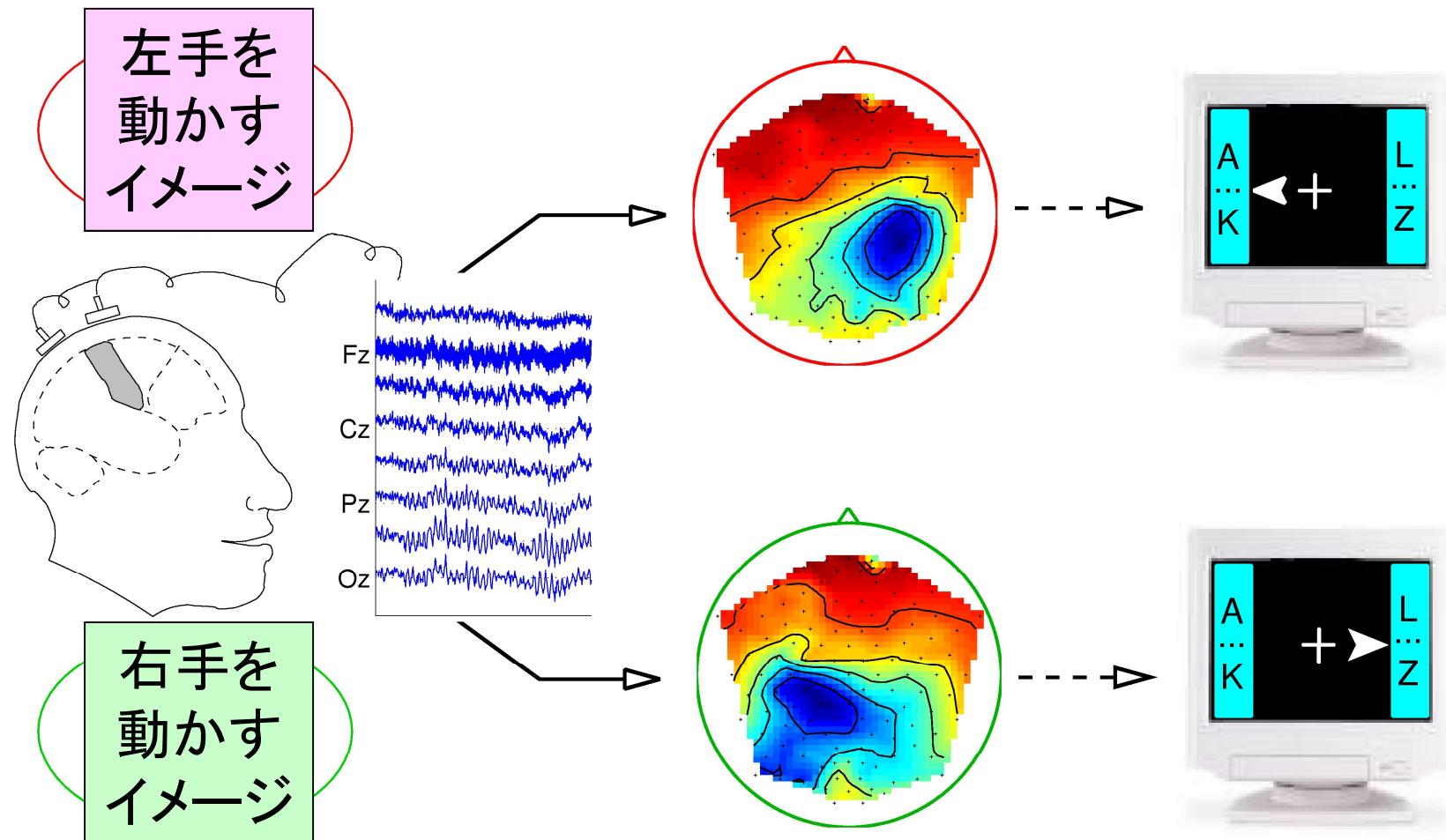
# ブレイン・コンピュータ インターフェース (BCI)

- 人間の**脳波**を読み取り、それに基づいてコンピュータを動かす
- 手足が動かない患者でもコンピュータが使えるようになる。



画像提供: フラウンホーファー研究所

# BCIの仕組み



画像提供: フラウンホーファー研究所

- 手を動かすイメージをするだけで、実際には動かさない

# コンピュータの学習

37

- 指示に従って、手を動かすイメージをする。

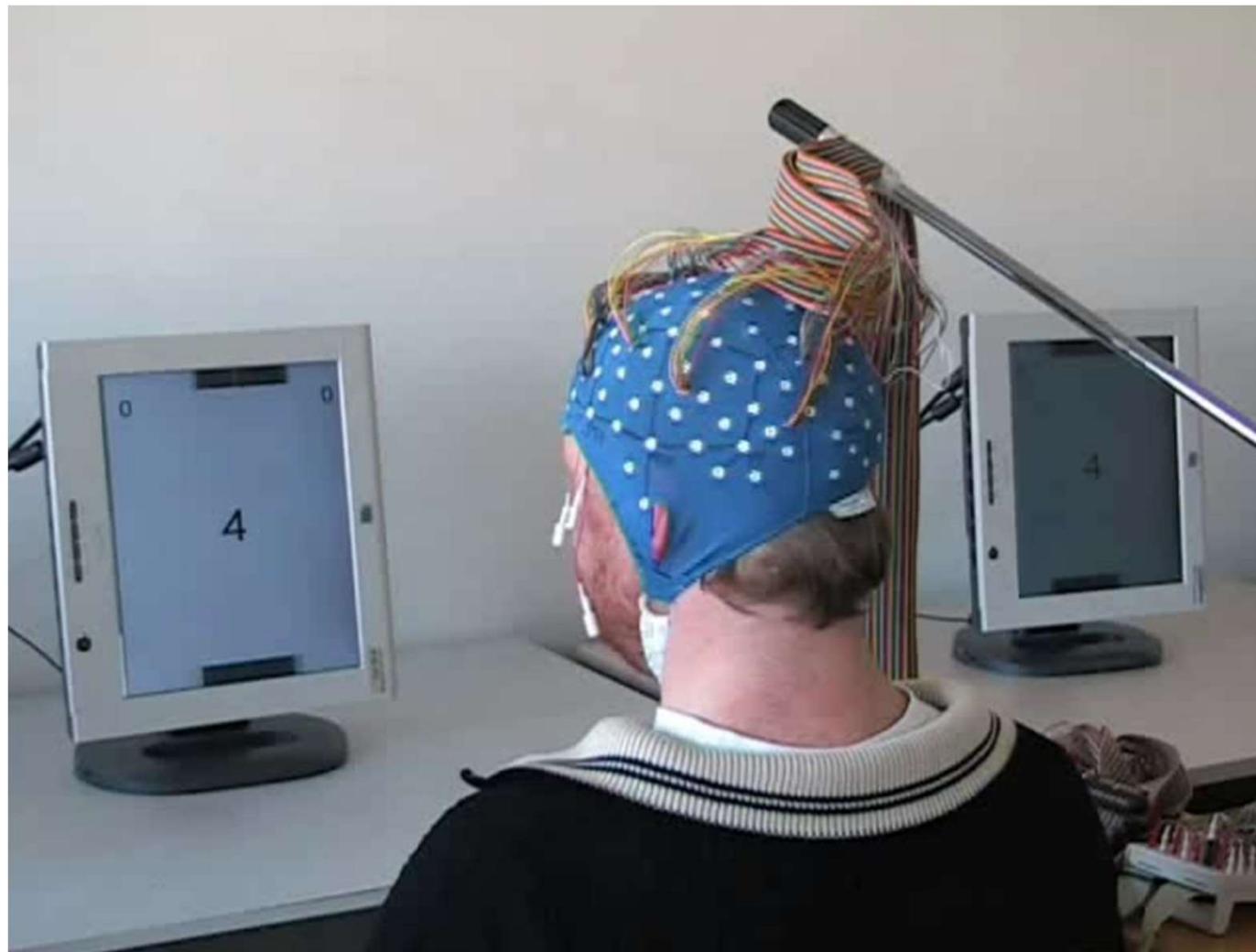


動画提供：フランツホーファー研究所

# ブレイン・ポン！

38

- 学習したコンピュータを使ってゲームをする



動画提供：フランツホーファー研究所



# 3種類の機械学習

39

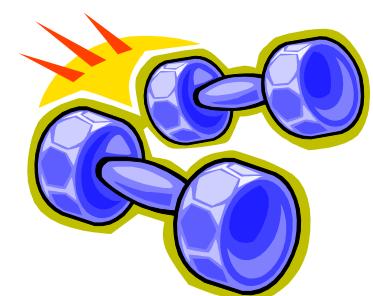
■ 教師付き学習：人間が教師となり、コンピュータの学習を手伝う



■ 教師なし学習：コンピュータが人間の手を介さずに自発的に学習する



■ 強化学習：ロボットが人間の手を介さずに自発的に意思決定を行う



# 教師なし学習

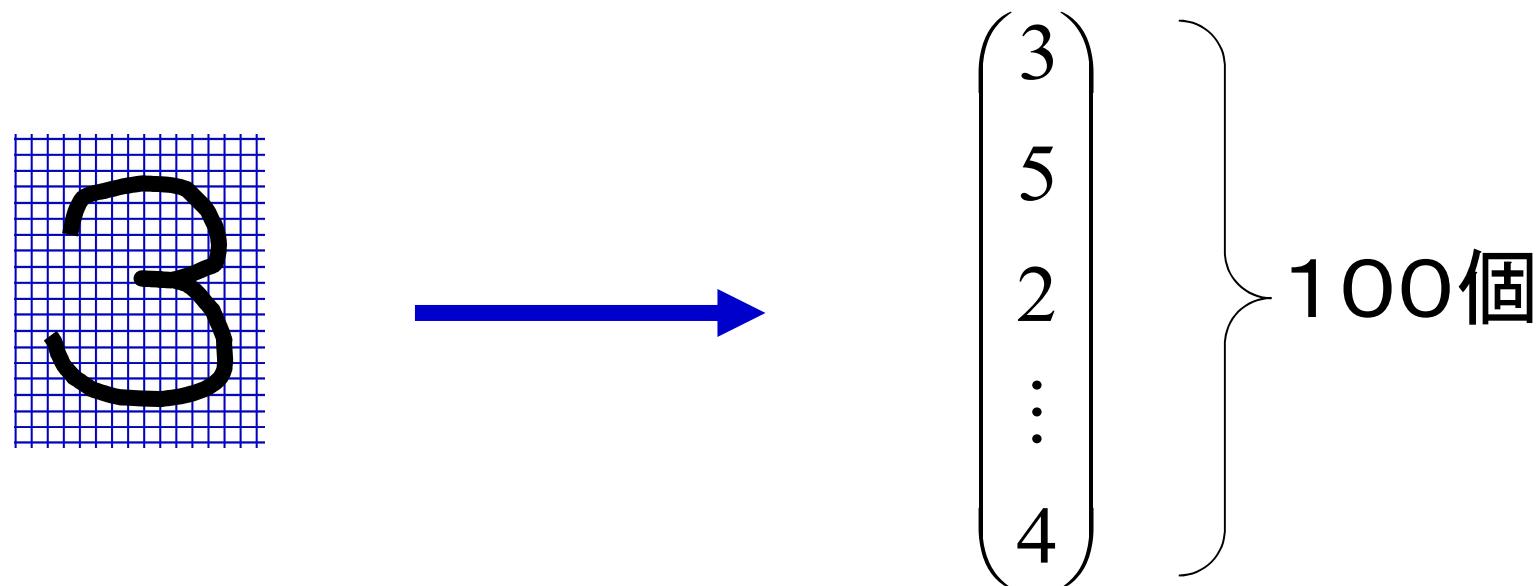
- 教師はない
- 自分ひとりで勉強する
- ゴールは勉強した知識から、  
**「有用なこと」**を見つけ出すこと



# 高次元データの可視化(1)

41

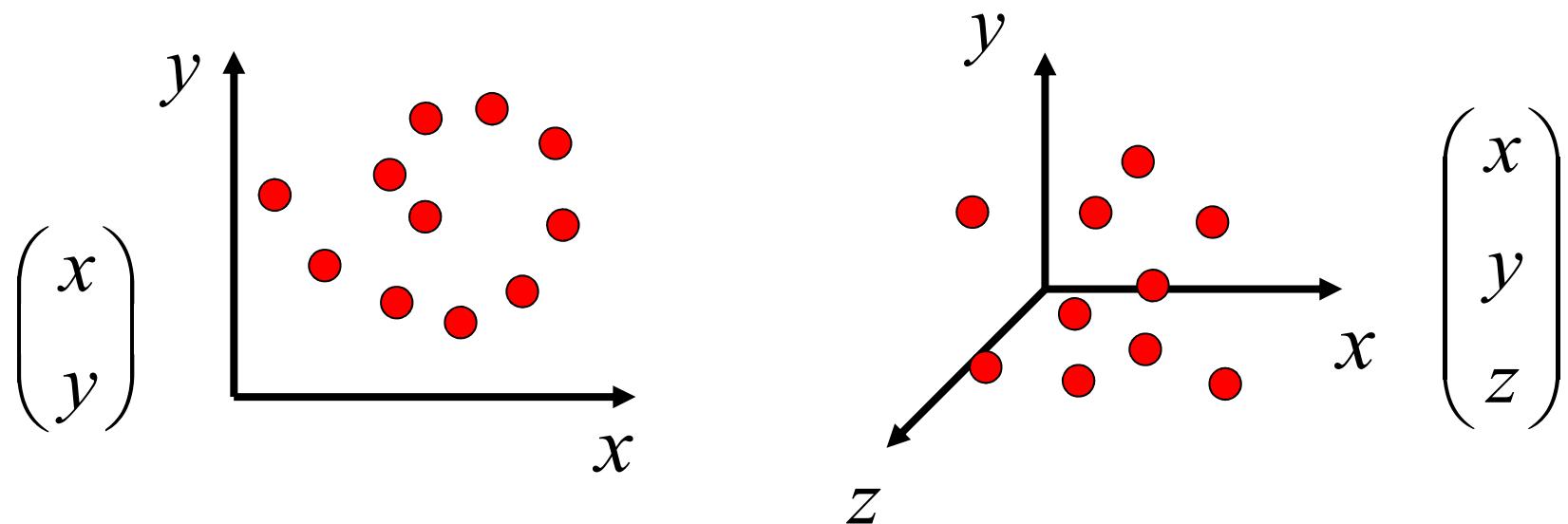
- ひとつの手書き数字が $10 \times 10$ 画素で構成されているとする。
- ベクトルにすると100次元



# 高次元データの可視化(2)

42

- 2次元や3次元のベクトルはグラフに図示することができる

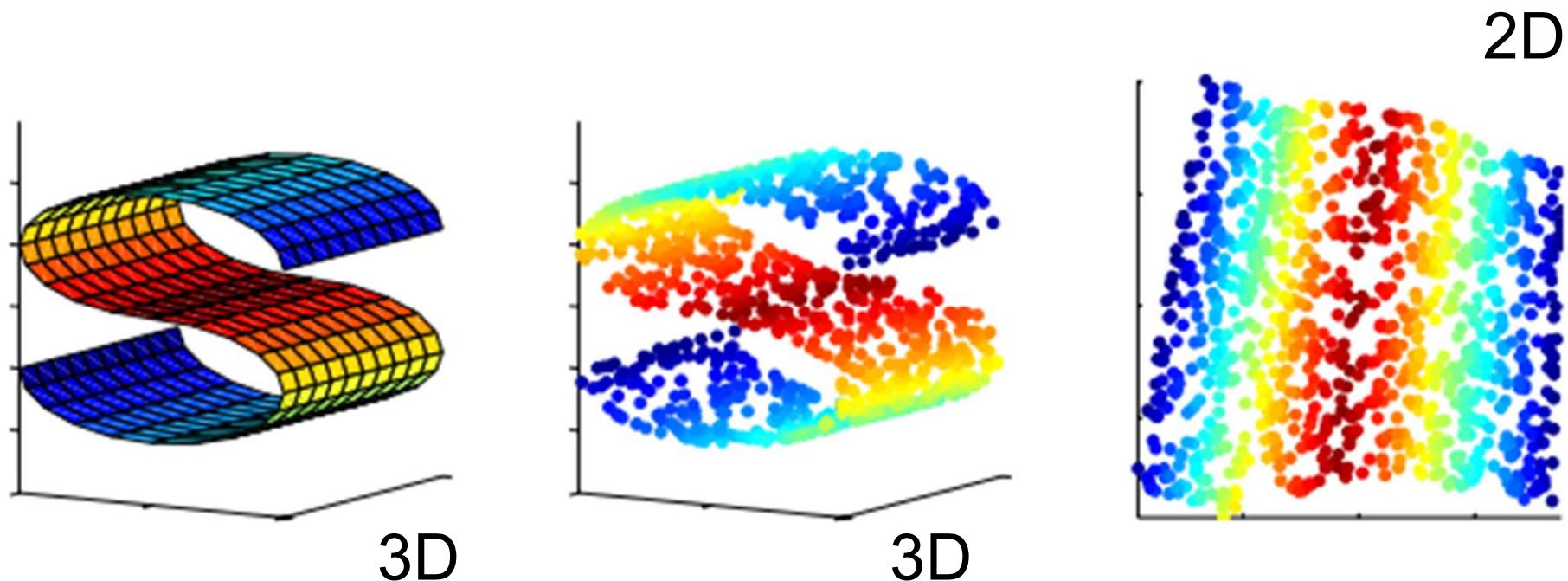


- 100次元のベクトルは図示できない！

# 高次元データの可視化(3)

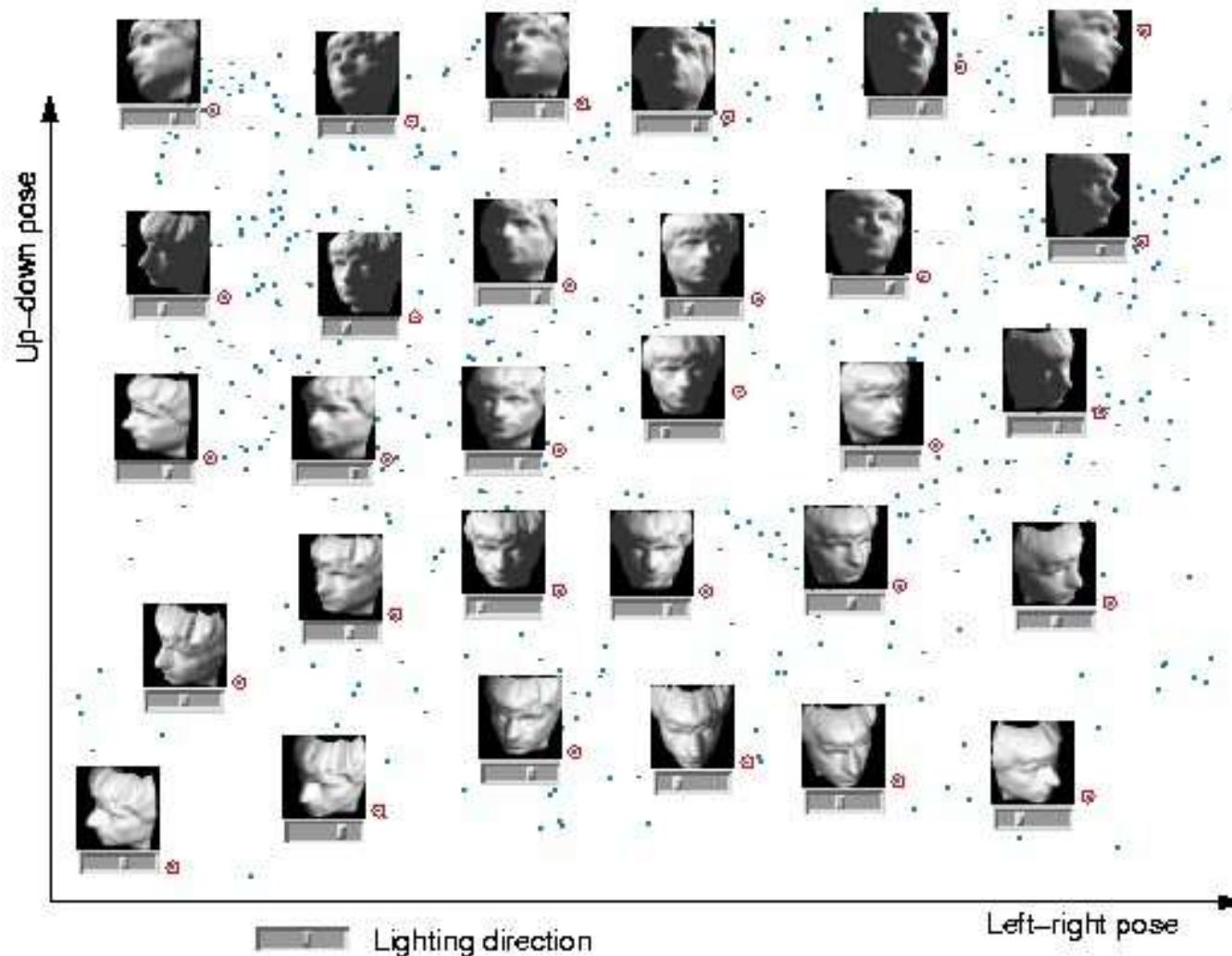
43

■ ゴール: もとの高次元のデータの特徴を残しつつ、低い次元(通常は2か3)に変換する



# 例

## ■ 顔の画像の分布( $64 \times 64 = 4096$ 次元)

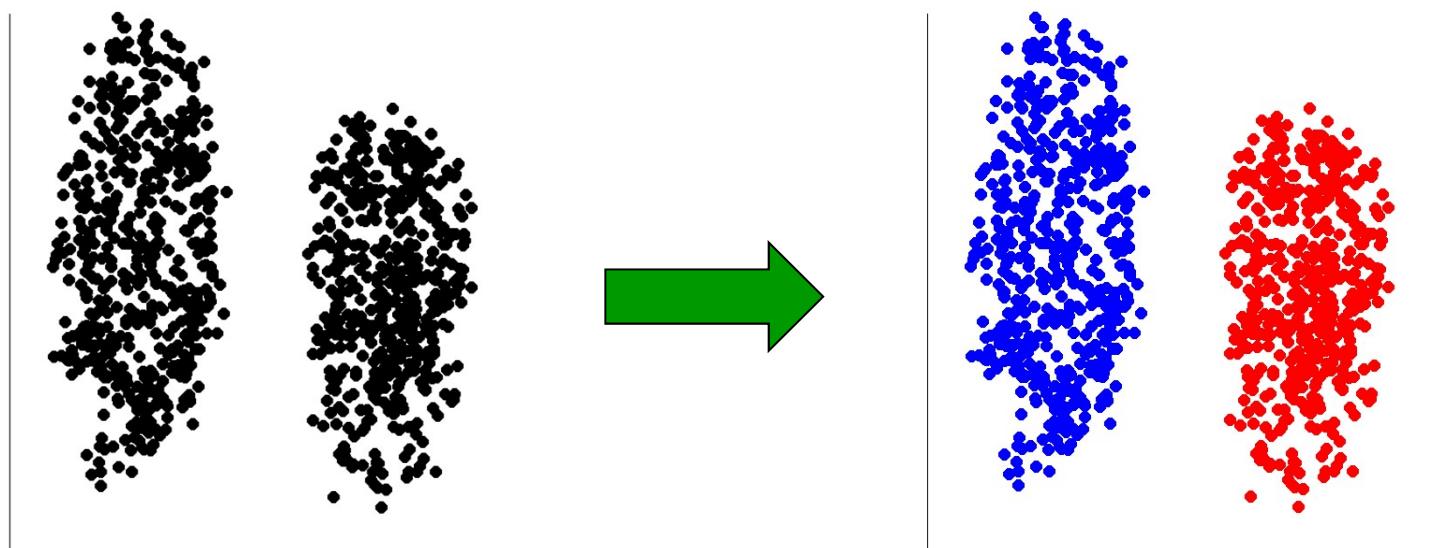


# データのクラスタリング

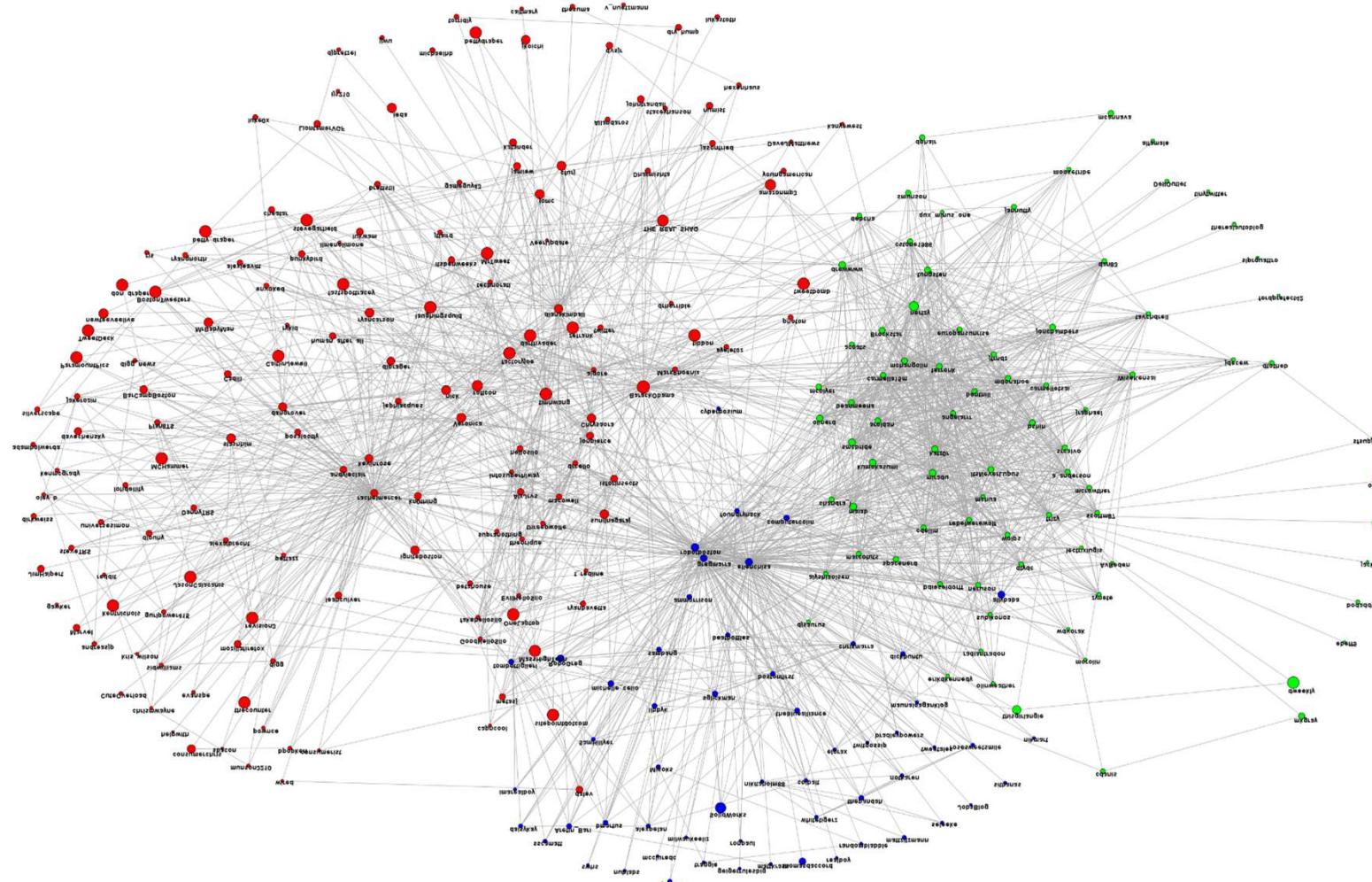
45

■ ゴール: 与えられたデータをいくつかのグループに分類したい

- 似たものは同じグループに
- 違うものは別のグループに



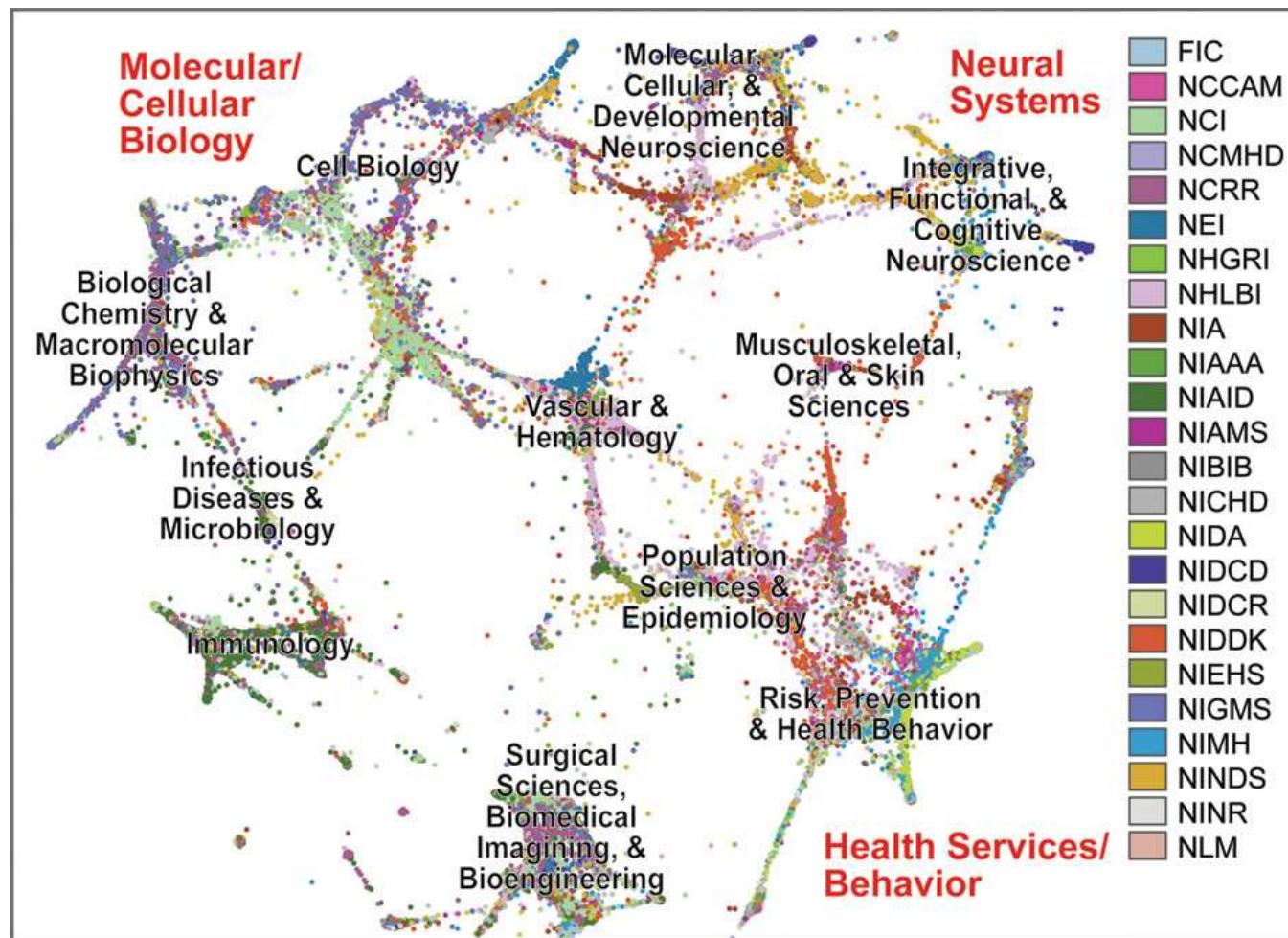
# ツイッターのフォロー関係からの コミュニティ発見 46



<http://ca.olin.edu/2008/realboy/images/world.png>

# アメリカ国立衛生研究所(NIH) の予算配分

47



<http://www.nature.com/nmeth/journal/v8/n6/images/nmeth.1619-F1.jpg>

# 異常検知

- ゴール: システムの異常を検知したい
- 通常と「質」の異なるデータが現れたら警報を鳴らす
  - コンピュータネットワークの侵入検知
  - 人工衛星の異常検知





# 3種類の機械学習

49

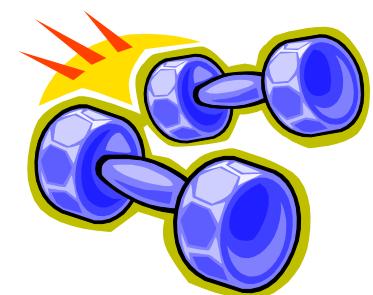
■ 教師付き学習：人間が教師となり、コンピュータの学習を手伝う



■ 教師なし学習：コンピュータが人間の手を介さずに自発的に学習する



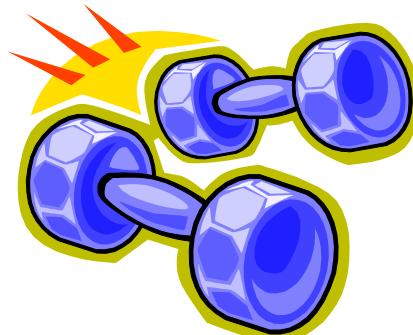
■ 強化学習：ロボットが人間の手を介さずに自発的に意思決定を行う



# 強化学習(1)

50

- 教師付き学習と同じく、教師の知識を学びたい
- しかし、教師は答えを教えてくれない
- 代わりに、生徒（ロボット）が予想した答えが正しければ**報酬**をくれる
- 生徒（ロボット）は、報酬が最大になるように学習する！



## 強化学習(2)

- 強化学習は赤ちゃんの学習と似ている
- 赤ちゃんは親にほめてもらえるような行動をとる



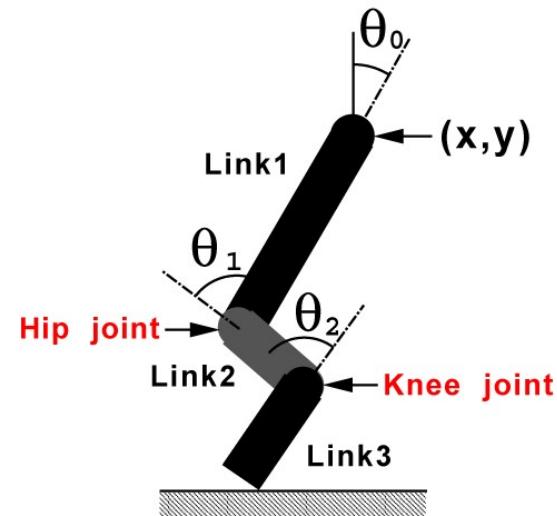
# ロボットの立ち上がり動作

## ■ ヒューマノイド(?)ロボット:

- 関節が二つ
- 関節の角度を操作できる

## ■ ゴール: 立ち上がり動作を学習させる

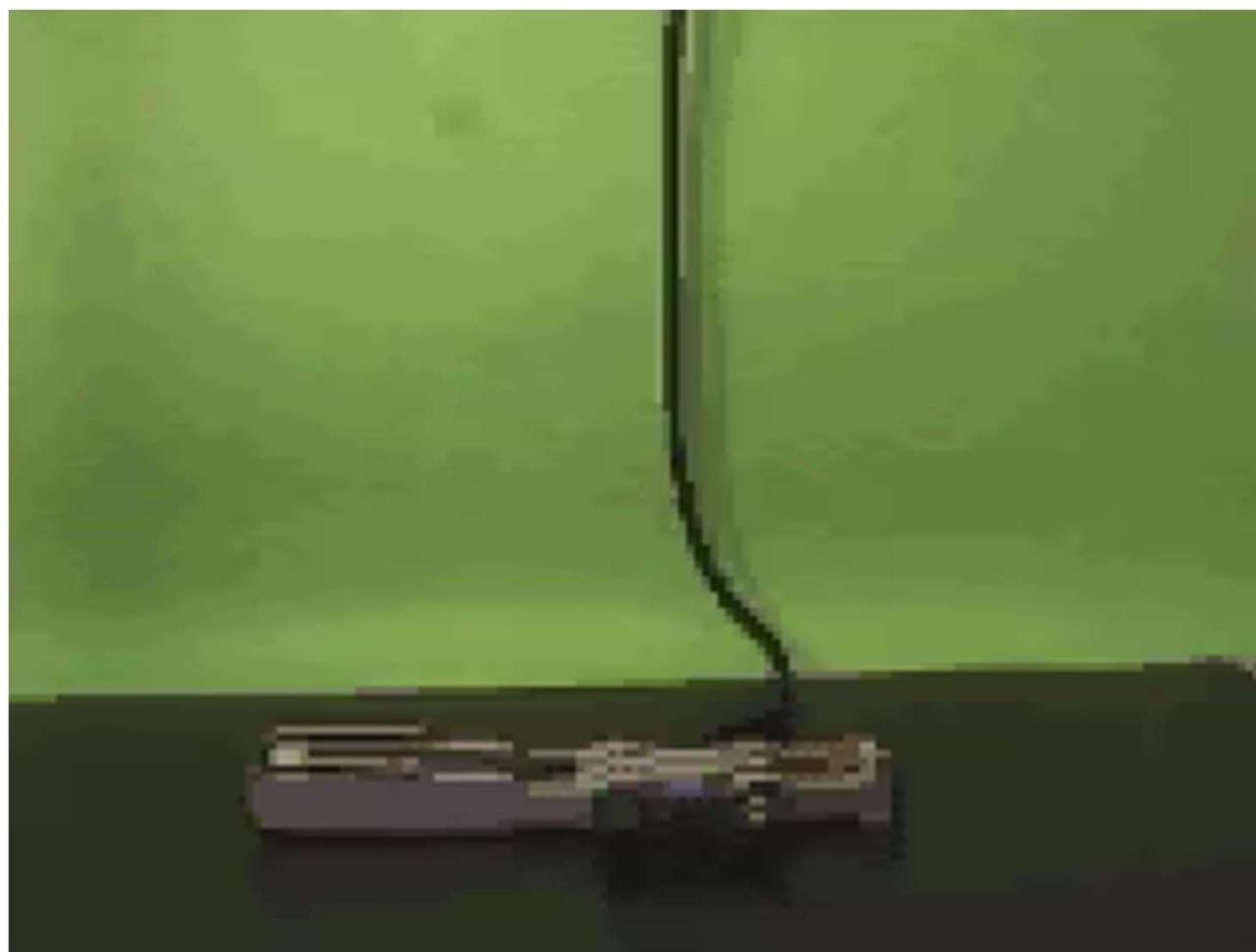
## ■ 報酬: 頭を高く持ち上げたときにプラスの報酬を与える



立ち上がり=頭の位置が最高

# デモ(1)

## ■ 学習前



[http://www.kawato.jst.go.jp/xmorimo/stup\\_movie.html](http://www.kawato.jst.go.jp/xmorimo/stup_movie.html)

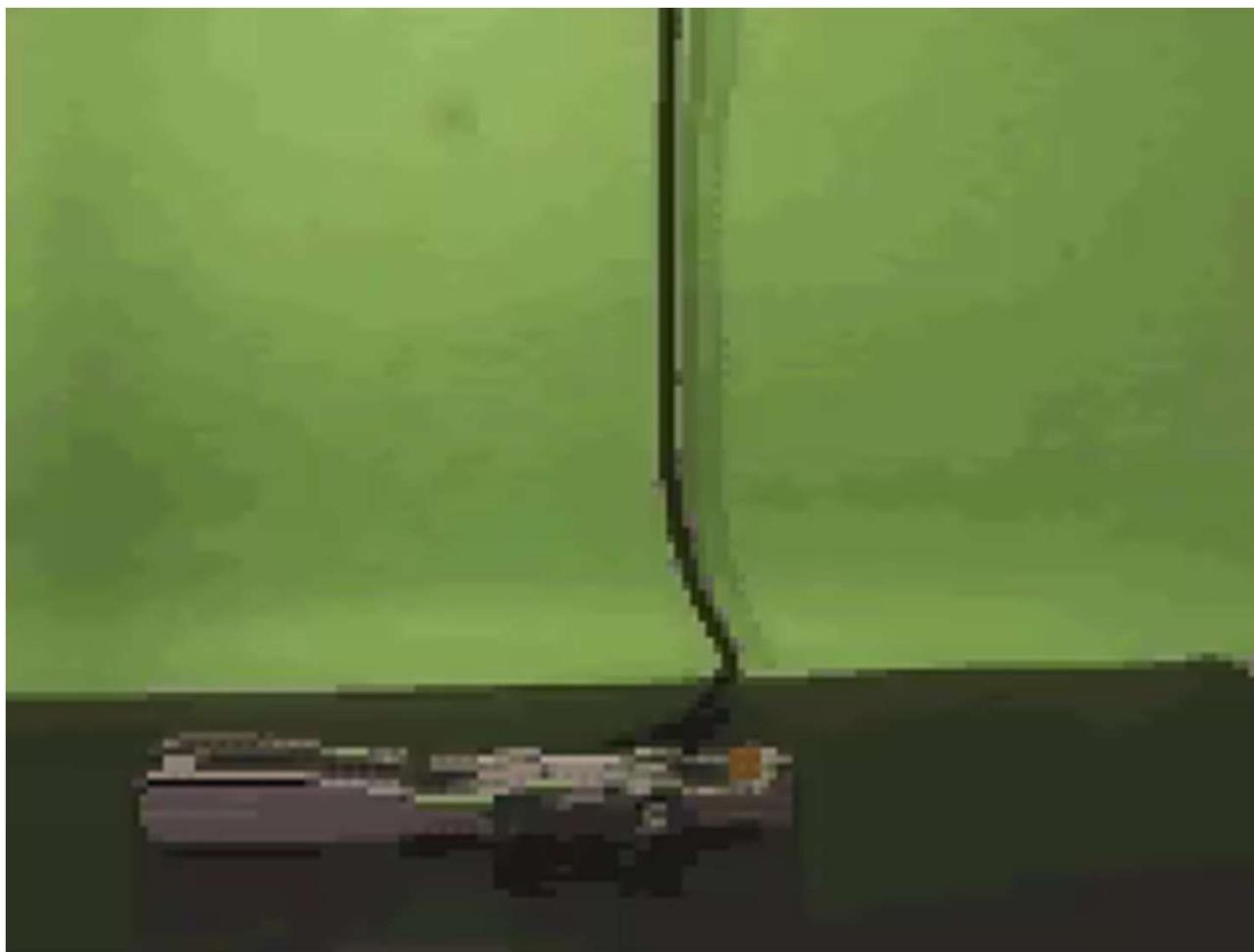
## デモ(2)

■ 750回の学習後



# デモ(3)

## ■ 920回の学習後



# バスケットボール

56





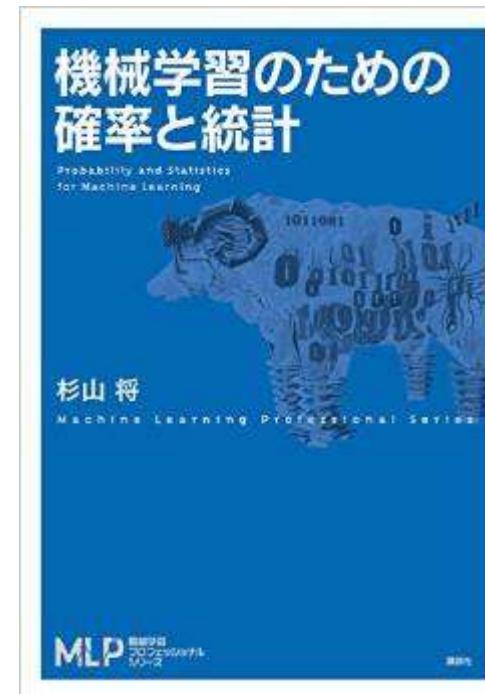
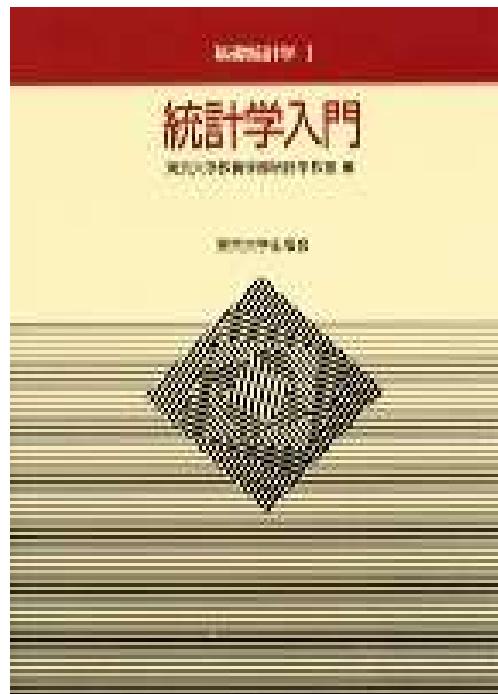
# 講義の流れ

57

1. 研究業界の動向
2. 機械学習の基礎
3. 確率・統計の基礎
  - A) 確率変数と確率分布
  - B) 多変数の確率分布
  - C) 確率変数の性質を表わす指標

# 参考書

- 東京大学教養学部統計学教室(編),  
統計学入門, 東京大学出版会, 1991年
- 杉山将, 機械学習のための確率と統計,  
講談社, 2015年



- 確率変数(random variable): とる値に対して確率が与えられている変数
- 実現値: 確率変数が実際にとる値
- 確率分布(probability distribution): 確率変数の実現値と確率との関係を関数として表現したもの
- 確率変数は大文字で、実現値は小文字で表わすことが多い
- カテゴリ  $y$ : 離散型の確率変数
- パターン  $x$ : 連続型の確率変数

# 離散型の確率変数と確率関数

60

- 离散型(discrete type)確率変数: 可算集合の中の値をとる確率変数
- 离散型の確率変数の確率分布: 確率変数がそれぞれの値をとる確率

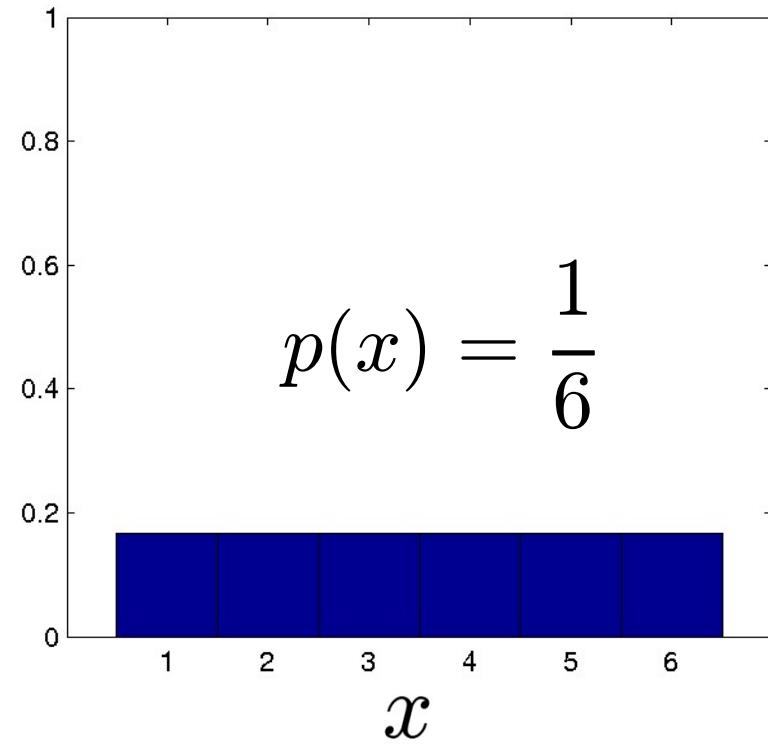
$$\text{Prob}(Y = y) = p(y)$$

$p(y)$ : 確率質量関数(probability mass function)

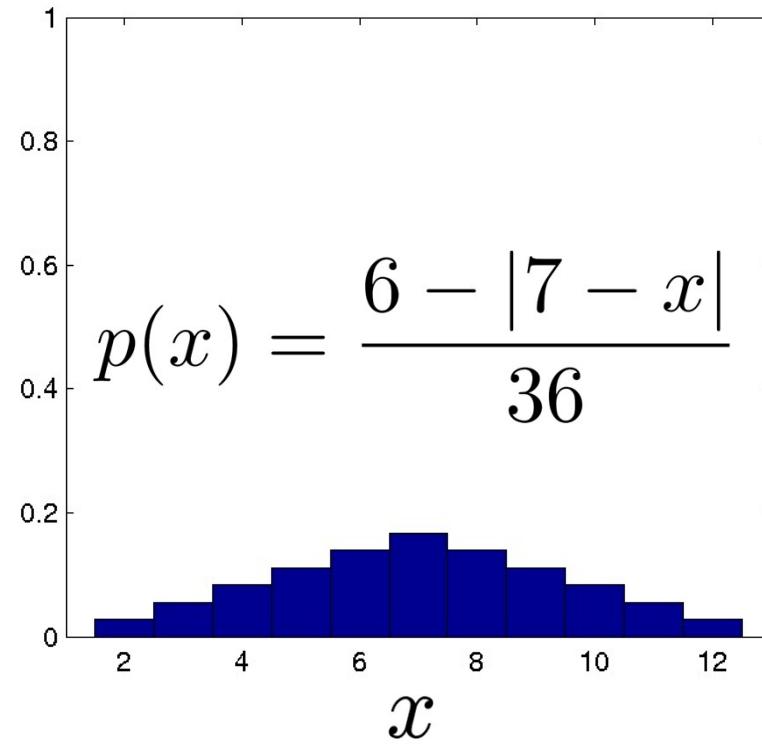
$$p(y) \geq 0, \quad \sum_y p(y) = 1$$

# 離散型の確率分布の例

61



さいころの出る目

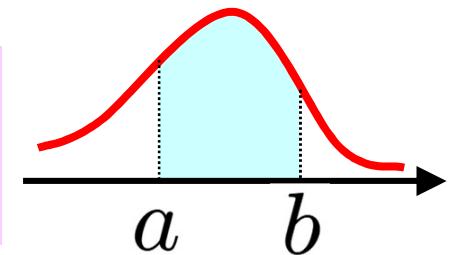


二つのさいころの  
出る目の和

# 連続型の確率変数と確率密度関数<sup>62</sup>

- 連続型(continuous type)確率変数: 連続値をとる確率変数
- 連続型の確率変数の確率分布: 確率変数が  $a$  以上  $b$  以下の値をとる確率

$$\text{Prob}(a \leq X \leq b) = \int_a^b p(x)dx$$



$p(x)$ : 確率密度関数(probability density function)

$$p(x) \geq 0, \quad \int p(x)dx = 1$$

- 注意: 連続型の確率変数がある値  $a$  をとる確率はゼロ!

$$\text{Prob}(X = a) = \int_a^a p(x)dx = 0$$



# 講義の流れ

63

1. 研究業界の動向
2. 機械学習の基礎
3. 確率・統計の基礎
  - A) 確率変数と確率分布
  - B) 多変数の確率分布
  - C) 確率変数の性質を表わす指標

# 同時確率

- $p(x, y) : x$  と  $y$  の同時確率(joint probability)
- 周辺化(marginalization) :

$$\sum_{y=1}^c p(x, y) = p(x)$$

$$\int_D p(x, y) dx = p(y)$$



周辺確率 (marginal probability)

- 独立性:  $x$  と  $x'$  が独立(independent) :

$$p(x, x') = p(x)p(x')$$

# 条件付き確率

■  $p(x | y)$  : 条件付き確率(conditional probability)

- $y$  が起こったもとで  $x$  が起こる確率

$$p(x|y) = \frac{p(x,y)}{p(y)} \text{ for } p(y) \neq 0$$

$$p(x) = \sum_y p(x,y)$$

■ 同時確率との関係 :

$$p(y | x)p(x) = p(x,y) = p(x | y)p(y)$$

# ベイズの定理

■ ベイズの定理(Bayes' theorem) :

$$p(y | x) = \frac{p(x | y)p(y)}{p(x)}$$

■ 事前確率(a priori probability)  $p(y)$  :

パターンを知る前のカテゴリの出現確率

■ 事後確率(a posteriori probability)  $p(y | x)$  :

パターンを知った後のカテゴリの出現確率

# ベイズの定理の使用例

67

## ■ 結構精度の良いウソ発見器：

- $X$ : 被験者の発言がウソか本当か
- $Y$ : ウソ発見器の出力
- ウソをウソと判定する確率：

$$p(Y=\text{ウソ} \mid X=\text{ウソ})=0.99$$

- 本当を本当と判定する確率：

$$p(Y=\text{本当} \mid X=\text{本当})=0.95$$

## ■ 被験者がウソをつく確率が

$$p(X=\text{ウソ})=0.001$$

のとき、ウソ発見器がウソだと言ったら信用すべきか？

# ベイズの定理の使用例

68

## ■ ベイズの定理を用いると

$$p(X=\text{ウソ}|Y=\text{ウソ}) = p(Y=\text{ウソ}|X=\text{ウソ})p(X=\text{ウソ})/p(Y=\text{ウソ})$$

- $p(Y=\text{ウソ}|X=\text{ウソ})=0.99$

- $p(X=\text{ウソ})=0.001$

- $P(Y=\text{ウソ})=p(Y=\text{ウソ}|X=\text{ウソ})p(X=\text{ウソ})$   
 $+p(Y=\text{ウソ}|X=\text{本当})p(X=\text{本当})$   
 $=p(Y=\text{ウソ}|X=\text{ウソ})p(X=\text{ウソ})$

$$\begin{aligned}&+ \{1-p(Y=\text{本当}|X=\text{本当})\} \{1-p(X=\text{ウソ})\} \\&= 0.99 \times 0.001 + \{1-0.95\} \times \{1-0.001\} \doteq 0.051\end{aligned}$$

$$\rightarrow p(X=\text{ウソ}|Y=\text{ウソ}) \doteq 0.99 \times 0.001 / 0.051 \doteq 0.019$$

$$\rightarrow p(X=\text{本当}|Y=\text{ウソ})=1-p(X=\text{ウソ}|Y=\text{ウソ}) \doteq 0.981$$

# ベイズの定理の使用例

$$p(X=\text{ウソ} | Y=\text{ウソ}) \doteq 0.019$$

$$p(X=\text{本当} | Y=\text{ウソ}) \doteq 0.981$$

- $p(X=\text{ウソ} | Y=\text{ウソ}) \ll p(X=\text{本当} | Y=\text{ウソ})$ なので、  
ウソ発見器の結果は信用すべきでない！
- ウソつきが少ない場合はウソ発見器は信用できない
- 実際,  $p(X=\text{ウソ} | Y=\text{ウソ}) > p(X=\text{本当} | Y=\text{ウソ})$ が  
成り立つためには,

$$p(X=\text{ウソ}) > 0.048$$

である必要がある



# 講義の流れ

70

1. 研究業界の動向
2. 機械学習の基礎
3. 確率・統計の基礎
  - A) 確率変数と確率分布
  - B) 多変数の確率分布
  - C) 確率変数の性質を表わす指標

# 確率変数の性質を表わす指標

71

- 期待値(expectation): 確率変数の値の平均  
(正確には確率による重み付きの平均)
- 確率変数  $X$  の期待値を  $E(X)$  で表す

- 離散型: 
$$E(Y) = \sum_y y p(y)$$

- 連続型: 
$$E(X) = \int x p(x) dx$$

# 期待値作用素の表記について

72

- $E(\cdot)$  は確率変数  $X$  に関する期待値を表す：

$$E(\cdot) = \int \cdot p(x)dx$$

- 正確には、 $E(\cdot)$  を  $E_X(\cdot)$  と表記すべきであるが、簡単のため省略している。

# 期待値演算の性質

- 定数は期待値をとっても値は変わらない:

$$E(c) = c$$

- 定数を足した期待値は、期待値に定数を足したものと等しい:

$$E(X + c) = E(X) + c$$

- 定数倍の期待値は、期待値の定数倍と等しい:

$$E(cX) = cE(X)$$

証明は宿題



期待値演算は線形

# 確率変数のばらつきの指標

- 期待値は確率変数の代表する値を表す指標
- 分散(variance): 確率変数の散らばり具合を表す指標

$$V(X) = E\{(X - E[X])^2\}$$

- 次式の方が計算しやすいこともある

$$\begin{aligned} V(X) &= E\{X^2 - 2XE[X] + (E[X])^2\} \\ &= E[X^2] - (E[X])^2 \end{aligned}$$

# 分散演算の性質

1. 定数の分散はゼロ

$$V(c) = 0$$

2. 定数を足したものの分散は、もとの分散と等しい

$$V(X + c) = V(X)$$

3. 定数倍の分散は、もとの分散に定数の2乗をかけたものと等しい

$$V(cX) = c^2 V(X)$$

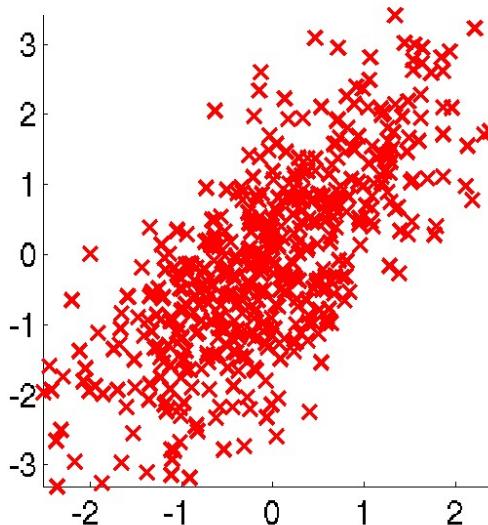
証明は宿題

# 共分散

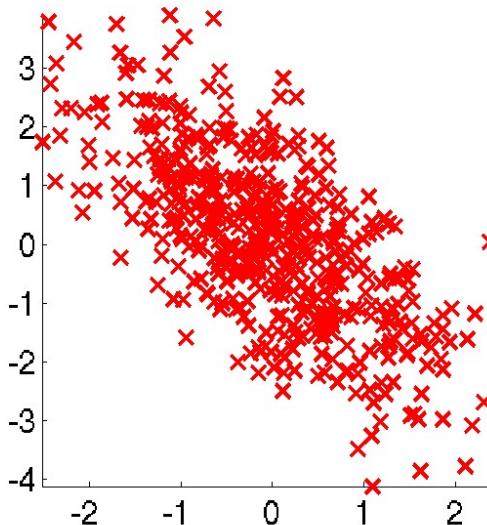
- $X$  と  $X'$  の共分散(covariance):

$$\text{Cov}(X, X') = E[(X - E[X])(X' - E[X'])]$$

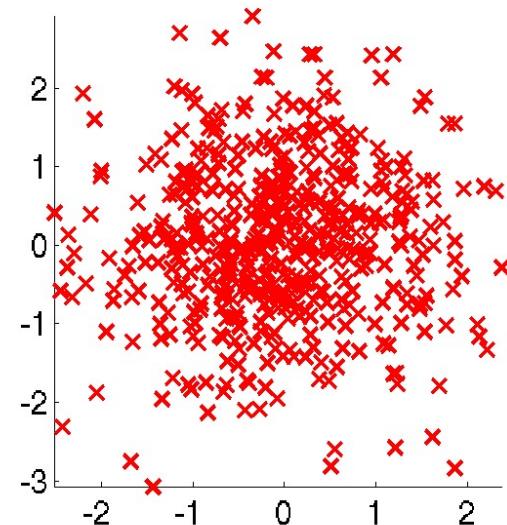
- $\text{Cov}(X, X') > 0$  のとき,  $X$  と  $X'$  の増減は同傾向
- $\text{Cov}(X, X') < 0$  のとき,  $X$  と  $X'$  の増減は逆傾向
- $\text{Cov}(X, X') = 0$  のとき,  $X$  と  $X'$  の増減は無関係



$$\text{Cov}(X, X') > 0$$



$$\text{Cov}(X, X') < 0$$



$$\text{Cov}(X, X') \not\simeq 0$$

# 分散と共分散の性質

77

1. 二つの確率変数  $X$  と  $X'$  の和の期待値は、それぞれの期待値の和と等しい：

$$E(X + X') = E(X) + E(X')$$

2. しかし  $X$  と  $X'$  の和の分散は、一般にはそれぞれの分散の和とは等しくない。

$$V(X + X') = V(X) + V(X') + 2\text{Cov}(X, X')$$

$$\text{Cov}(X, X') = E[(X - E[X])(X' - E[X'])]$$

証明は宿題

# 共分散の使用法の例

$$V(X + X') = V(X) + V(X') + 2\text{Cov}(X, X')$$

- A社の株価を  $X$ , B社の株価を  $X'$  とする
- $\text{Cov}(X, X') > 0$  のとき,
 
$$V(X + X') > V(X) + V(X')$$
  - A, B両社の株を買うと, 分散が拡大
  - 变動リスクが増大し, 資産価値は不安定

- $\text{Cov}(X, X') < 0$  のとき,
 
$$V(X + X') < V(X) + V(X')$$
  - A, B両社の株を買うと, 分散が縮小
  - 变動リスクが抑制され、資産価値は安定

# 分散共分散行列

■ 分散共分散行列(variance-covariance matrix):

$$\Sigma = E \left[ \left\{ \begin{pmatrix} X \\ X' \end{pmatrix} - E \begin{pmatrix} X \\ X' \end{pmatrix} \right\} \left\{ \begin{pmatrix} X \\ X' \end{pmatrix} - E \begin{pmatrix} X \\ X' \end{pmatrix} \right\}^\top \right]$$

$$= \begin{pmatrix} V(X) & \text{Cov}(X, X') \\ \text{Cov}(X, X') & V(X') \end{pmatrix}$$

■ 対角成分は分散, 非対角成分は共分散.

# 相関

- 相関係数(correlation coefficient): 共分散を標準偏差で割った値

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

- 相関係数は  $-1 \leq \rho_{XY} \leq 1$  を満たす.
- $\rho_{XY} > 0$  のとき, 正の相関があるという.
- $\rho_{XY} < 0$  のとき, 負の相関があるという.
- $\rho_{XY} = 0$  のとき, 無相関(uncorrelated)であるという.
- $|\rho_{XY}| \approx 1$  のとき,  $X$  と  $Y$  の増減関係はより確定的になる.



# 講義の流れ

81

1. 研究業界の動向
2. 機械学習の基礎
3. 確率・統計の基礎

# 宿題1

■ 確率と統計の授業が好きな人, 嫌いな人の確率が

$$p(X=\text{好})=0.8 \quad p(X=\text{嫌})=0.2$$

であるとする. また, 確率と統計の授業が好きな人の中で授業中眠たい人, および, 確率と統計の授業が嫌いな人の中で授業中眠たい人の確率が, それぞれ

$$p(Y=\text{眠} \mid X=\text{好})=0.25 \quad p(Y=\text{眠} \mid X=\text{嫌})=0.25$$

であるとする.

- A)  $p(X=\text{好}, Y=\text{眠})$ を求めよ.
- B)  $p(Y=\text{眠})$ を求めよ.
- C)  $p(X=\text{好} \mid Y=\text{眠})$ を求めよ.
- D) 確率と統計の好き嫌いと授業中眠たい事は独立か?

## 宿題2

■ 以下を証明せよ:

A) 定数は期待値をとっても値は変わらない:

$$E(c) = c$$

B) 定数を足した期待値は、期待値に定数を足したものと等しい:

$$E(X + c) = E(X) + c$$

C) 定数倍の期待値は、期待値の定数倍と等しい:

$$E(cX) = cE(X)$$

# 宿題3

■ 以下を証明せよ:

A) 定数の分散はゼロ

$$V(c) = 0$$

B) 定数を足したものの分散は、もとの分散と等しい

$$V(X + c) = V(X)$$

C) 定数倍の分散は、もとの分散に定数の2乗をかけたものと等しい

$$V(cX) = c^2 V(X)$$

# 宿題4

■ 以下を証明せよ:

A) 二つの確率変数  $X$  と  $X'$  の和の期待値は、それぞれの期待値の和と等しい:

$$E(X + X') = E(X) + E(X')$$

B) しかし  $X$  と  $X'$  の和の分散は、一般にはそれぞれの分散の和とは等しくない。

$$V(X + X') = V(X) + V(X') + 2\text{Cov}(X, X')$$

$$\text{Cov}(X, X') = E[(X - E[X])(X' - E[X'])]$$

- 宿題は次回の講義までに、  
ITC-LMSにアップロードすること