

確率と統計(○) 「導入」

- 担当教員: 杉山 将 (計算工学専攻)
- 居室: W8E-406
- 電子メール: sugi@cs.titech.ac.jp
- 授業のウェブサイト:
<http://sugiyama-www.cs.titech.ac.jp/~sugi/>

成績評価

- 宿題: $s_H, \quad 0 \leq s_H \leq 60$
- テスト: $s_T, \quad 0 \leq s_T \leq 60$
- 演習: $s_E, \quad 0 \leq s_E (\leq 100)$
- 最終評価: $s_F = \min(100, s), \quad s = s_H + s_T + s_E$
- 但し, 最終段階で s の値を調整する可能性あり(例えば, ほとんどの人が単位を取れない場合など).
- 宿題が満点ならば, 試験を受けなくても単位が取れる.
- テストが満点ならば, 宿題を一度も出さなくても単位が取れる.
- 演習のときに, 前で発表した学生は成績アップ!

連絡事項

■ テキスト:

- 統計学入門, 東京大学出版会

■ その他参考書:

- 自然科学の統計学, 東京大学出版会
- 渡辺澄夫・村田昇, 確率と統計—情報学への架橋—, コロナ社
- 杉山将, 統計的機械学習, オーム社

統計学とは

■ 統計学(statistics)

データから現象の法則などの有用な情報を取り出すための方法論を議論する学問。

- 記述統計学(descriptive statistics)
- 数理統計学(mathematical statistics)

記述統計学

- 全てのデータを丹念に調べ、規則性などを見つけ出す.

- 国勢調査

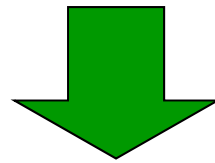
5年に一度、日本に居住するすべての人の性別、年齢、国籍、就業状態、仕事の種類などを調べる.

- 入試の成績

毎年、全員の試験の点数を調べる.

数理統計学

- 記述統計学はデータを整理・要約する方法論を議論する学問である.
- しかし、現実にはすべてのデータを採集するのが大変な場合や、原理的にデータを採取できない場合もある。



- 数理統計学(mathematical statistics)
確率論に基づいて、一部のデータから論理性のある推測で全体の法則を見出すための方法論を議論する。

数理統計学が役立つ例

- **全数調査**: 国勢調査のように全員の情報を調べるのは時間もお金もかかる.
- **破壊検査**: 携帯電話を1メートルの高さから落としたときに壊れる確率がどの程度か知りたいとき, 実際に全ての携帯電話を落として壊れる確率を調べることはできない.
- **予測**: 明日の降水量を知りたいとき, 明日にならないと明日の降水量のデータは得られない.

数理統計学が役立つ例（続き）

- **学習するコンピュータ**: 人間が幾つかの例を教えれば, あとはコンピュータが自分で考えて自動的に処理してくれる.
- 本日の講義では, 数理統計学を活用した学習するコンピュータの例を紹介する.

学習

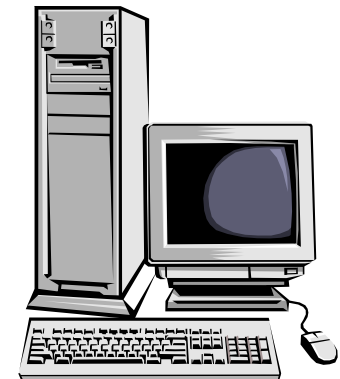
- 勉強のできる賢い人は何がすごい？
 - **暗記能力**：教えられたことをきちんと覚える
 - **応用能力**：教わっていないことも、知っている知識から類推する



最近のコンピュータはすごい！

10

- 近年、パソコンの性能は飛躍的に向上
 - メモリ: 数ギガバイト
 - ハードディスク: 数百ギガバイト
 - CPUクロック: 数ギガヘルツ
- たくさんのコンピュータをつなげれば、性能は更に向上
- 東工大のスパコン: TSUBAME



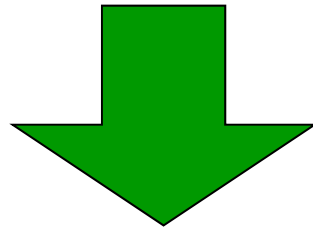
コンピュータは賢い？

■ 記憶力:

コンピュータ $>$ 人間

■ 応用力:

コンピュータ $<$ 人間



コンピュータに応用力をつけさせたい！

学習の種類

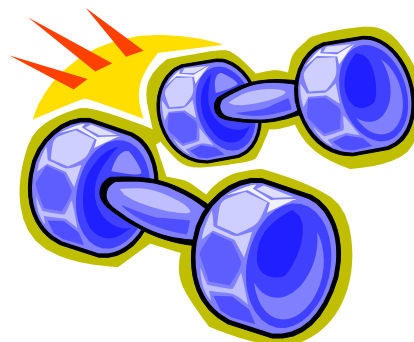
教師有り学習



教師無し学習

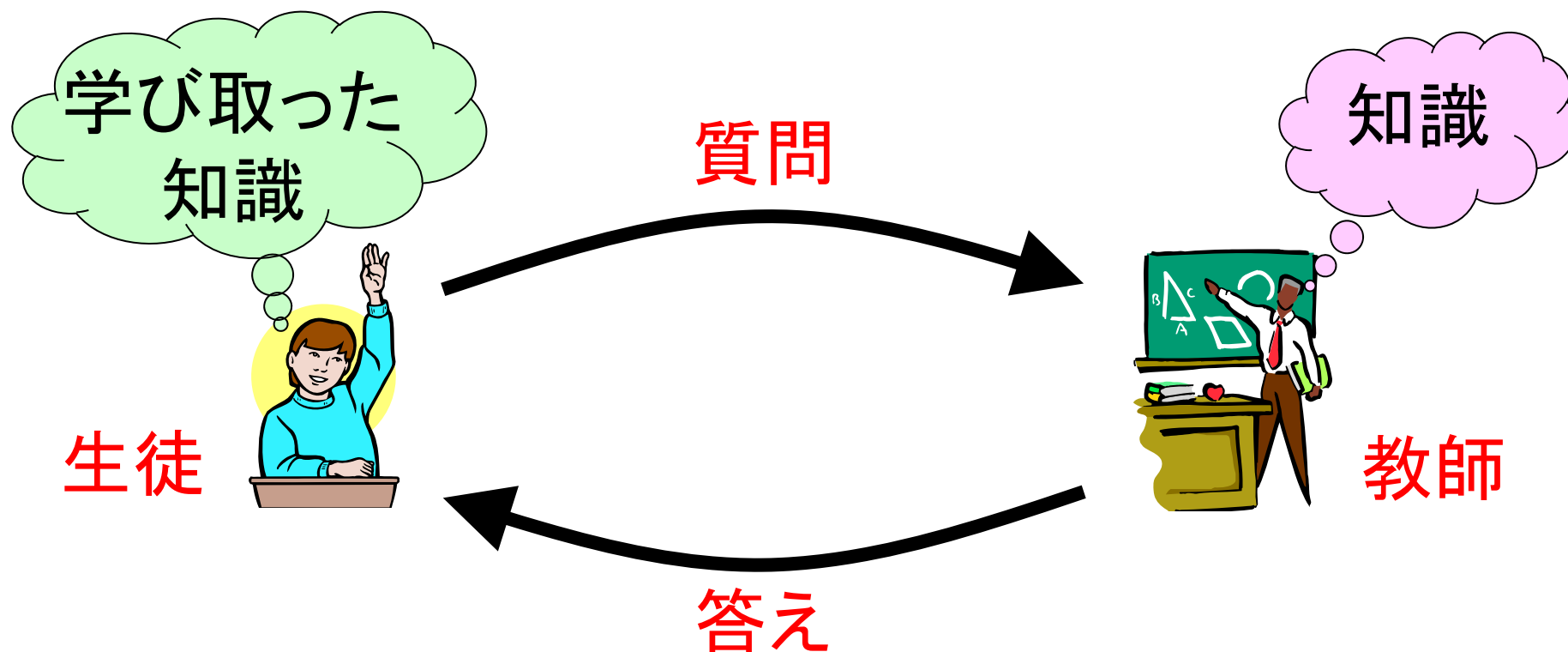


強化学習



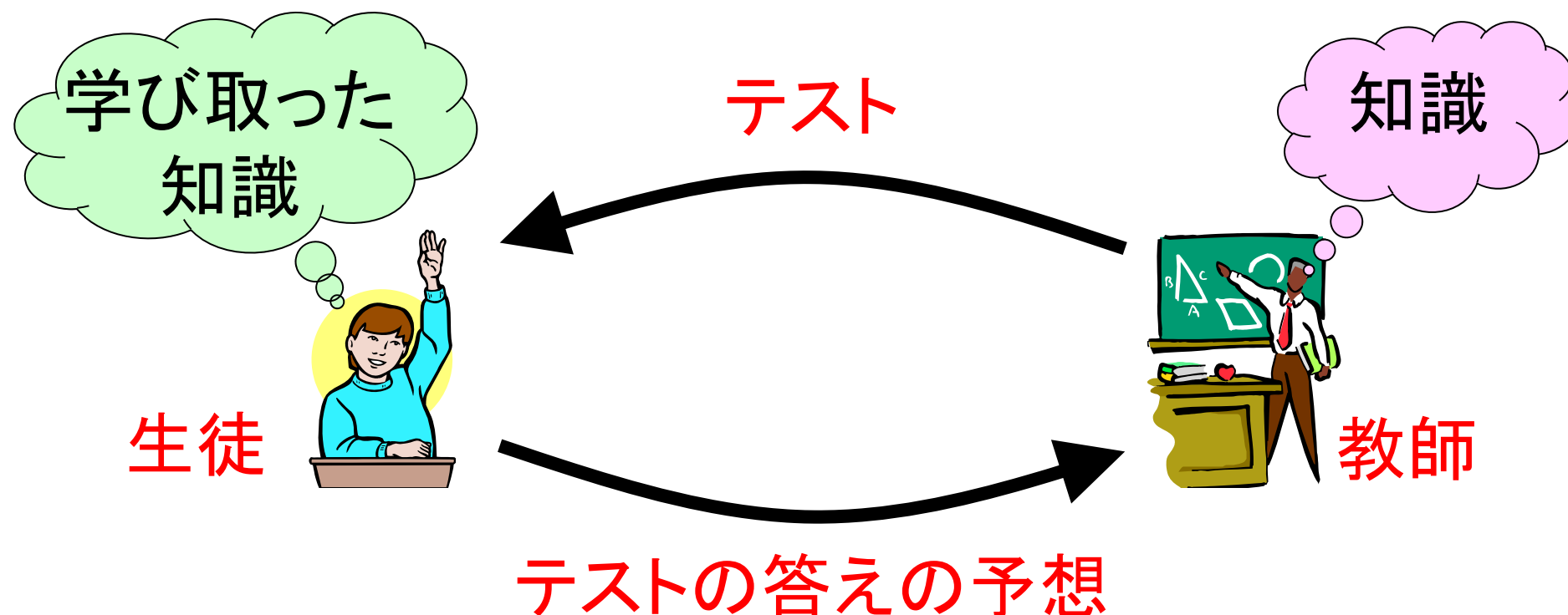
教師有り学習

- **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

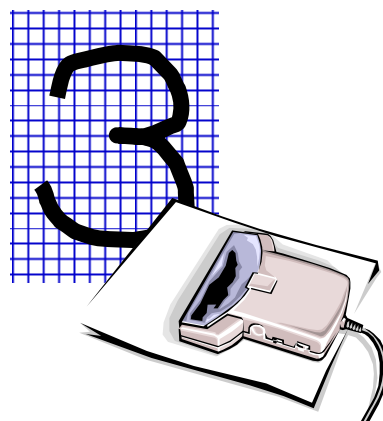
- 教わった答えを覚えるだけでなく、
教師の知識**全体**を学び取れば、
答えを教わっていない新しい質問
(テスト)に対しても正しく解答できる



例：手書き数字認識

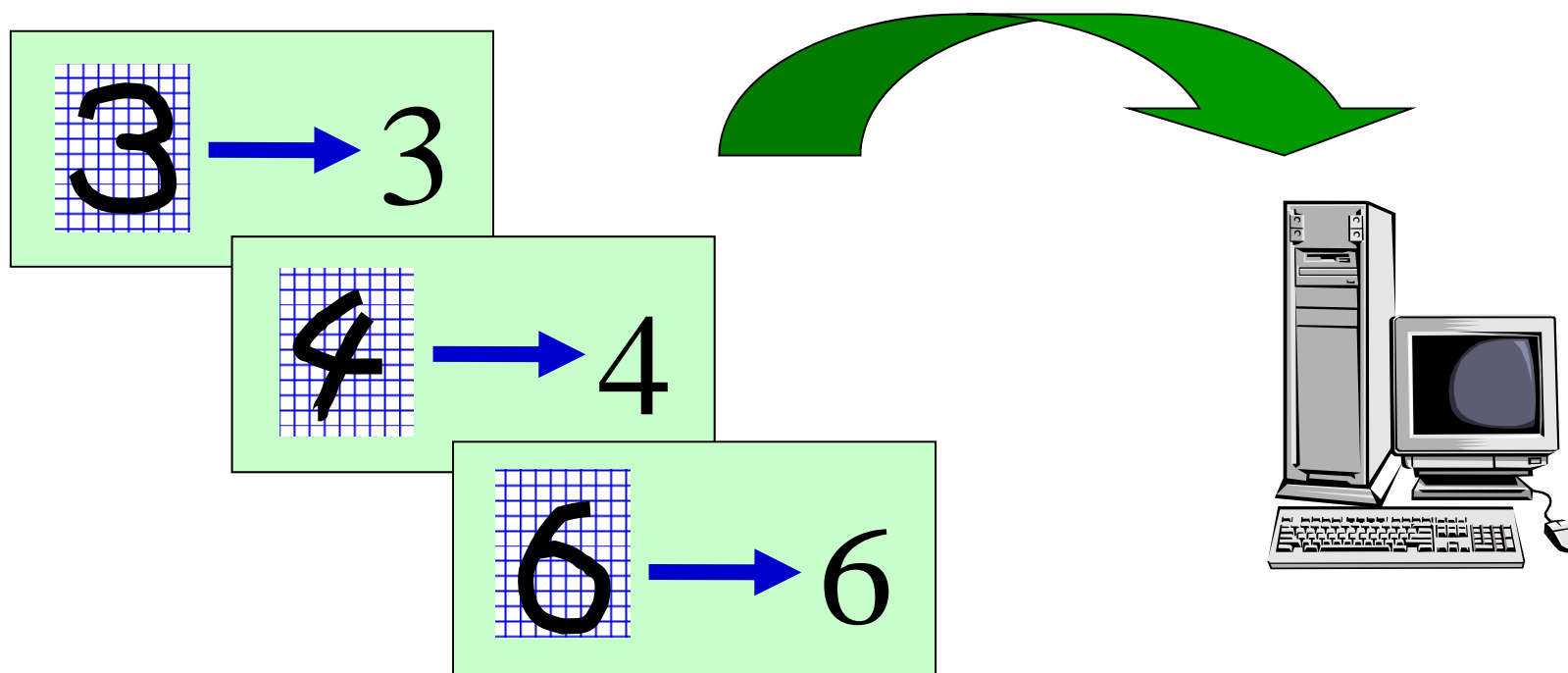
■ 手書き数字の認識

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



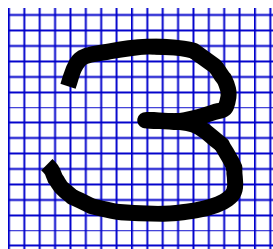
例：手書き数字認識（続き）

- 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 10×10 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！

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.

脳波信号生成の例

- 入力: 脳波
- 出力: 「右」, 「左」

- 「右」信号の生成: 右手を動かすイメージをする
- 「左」信号の生成: 左手を動かすイメージをする
- 注意
 - 実際に手足は動かさない！手, 足, 顔(目も含む)につけた筋電センサーが反応した場合は, そのデータを却下する
 - 脳内への電極の埋め込みなどは行なわない

学習用のデータ作成

ゲーム

対戦型ゲーム

学習の種類

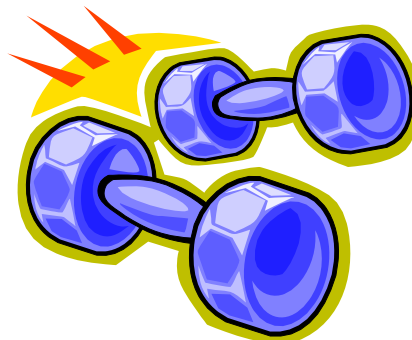
教師有り学習



教師無し学習



強化学習



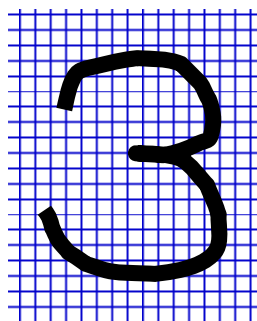
教師無し学習

- 教師はいない
- 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)

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

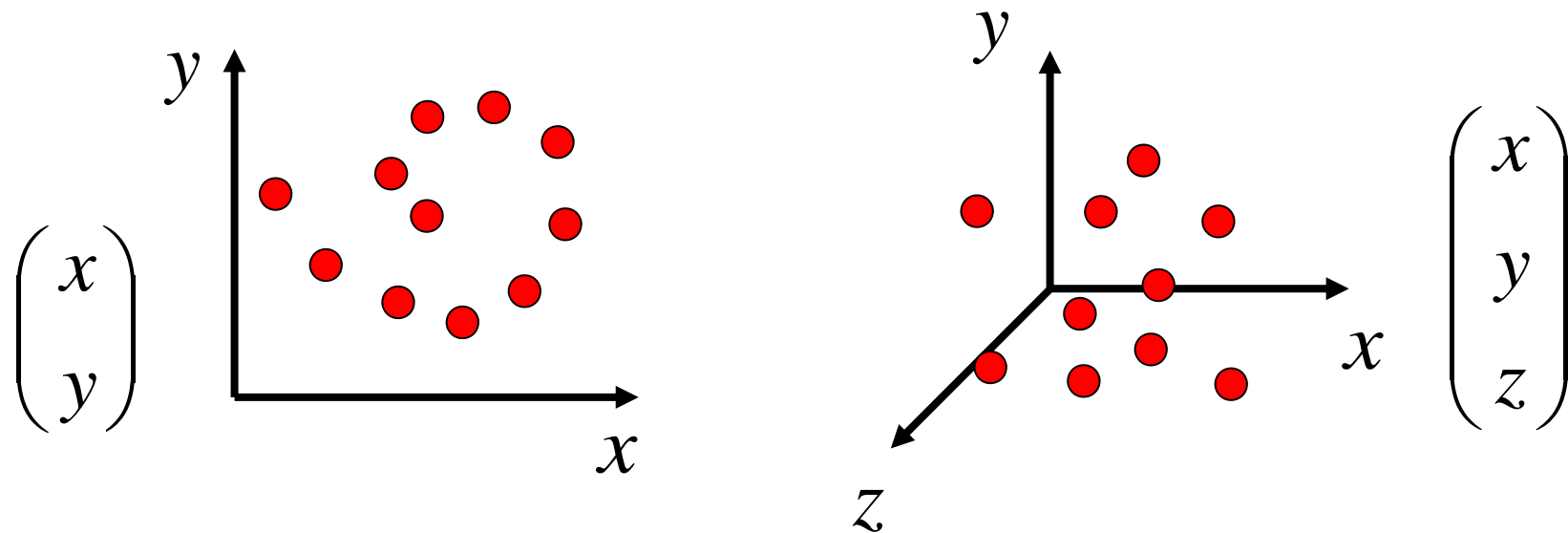


$$\begin{pmatrix} 3 \\ 5 \\ 2 \\ \vdots \\ 4 \end{pmatrix} \left. \vphantom{\begin{pmatrix} 3 \\ 5 \\ 2 \\ \vdots \\ 4 \end{pmatrix}} \right\} 100\text{個}$$

高次元データの可視化(2)

27

- 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)

28

- **ゴール**: もとの高次元のデータの特徴を残したまま, 低い次元(通常は2か3)に変換する

2D

3D

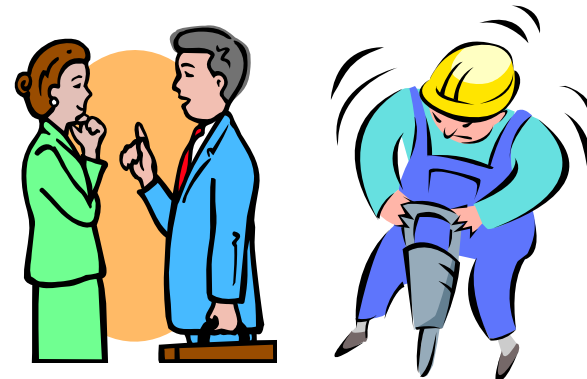
3D

例

- 顔の画像の分布 ($64 \times 64 = 4096$ 次元)

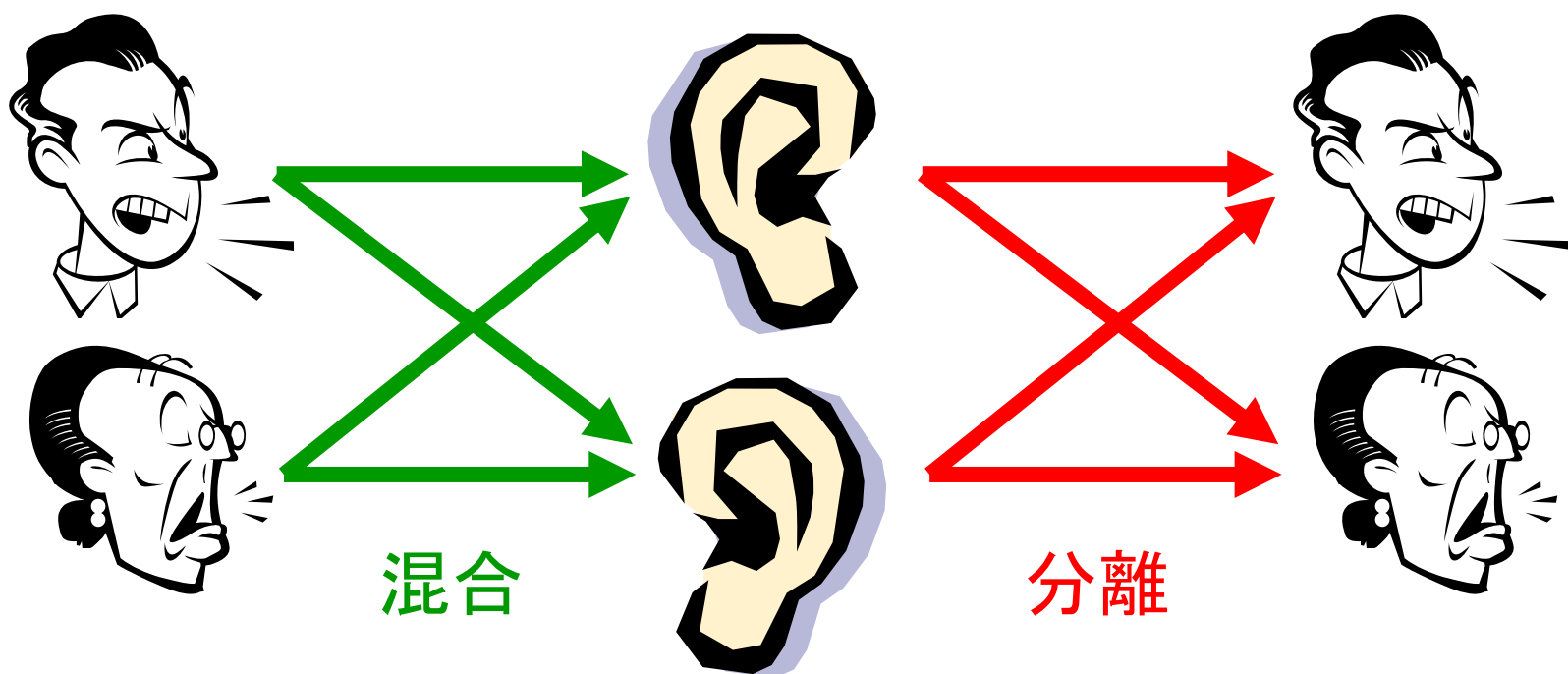
聖徳太子コンピュータ

- 聖徳太子は10人の会話を同時に理解したらしい
- 普通の人でも、2人の会話くらいなら理解できる？
- 工事現場で立ち話しをしても、話は聞き取れる



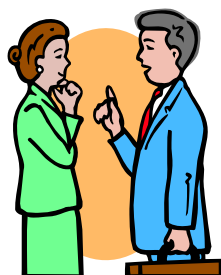
ブラインド信号源分離

- 二人の会話が混ざった音声信号を二つの耳で聞く.
- **ゴール**: それらをもとの二つの音声信号に分離したい.



例

	混ざった 信号	分離した 信号1	分離した 信号2
会話＋会話			
会話＋楽器			



学習の種類

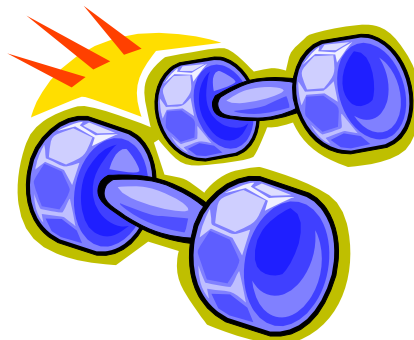
教師有り学習



教師無し学習

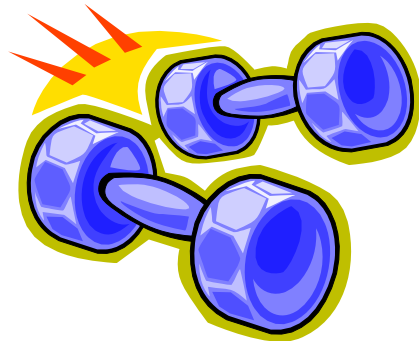


強化学習



強化学習(1)

- 教師有り学習と同じく，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いければ**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)

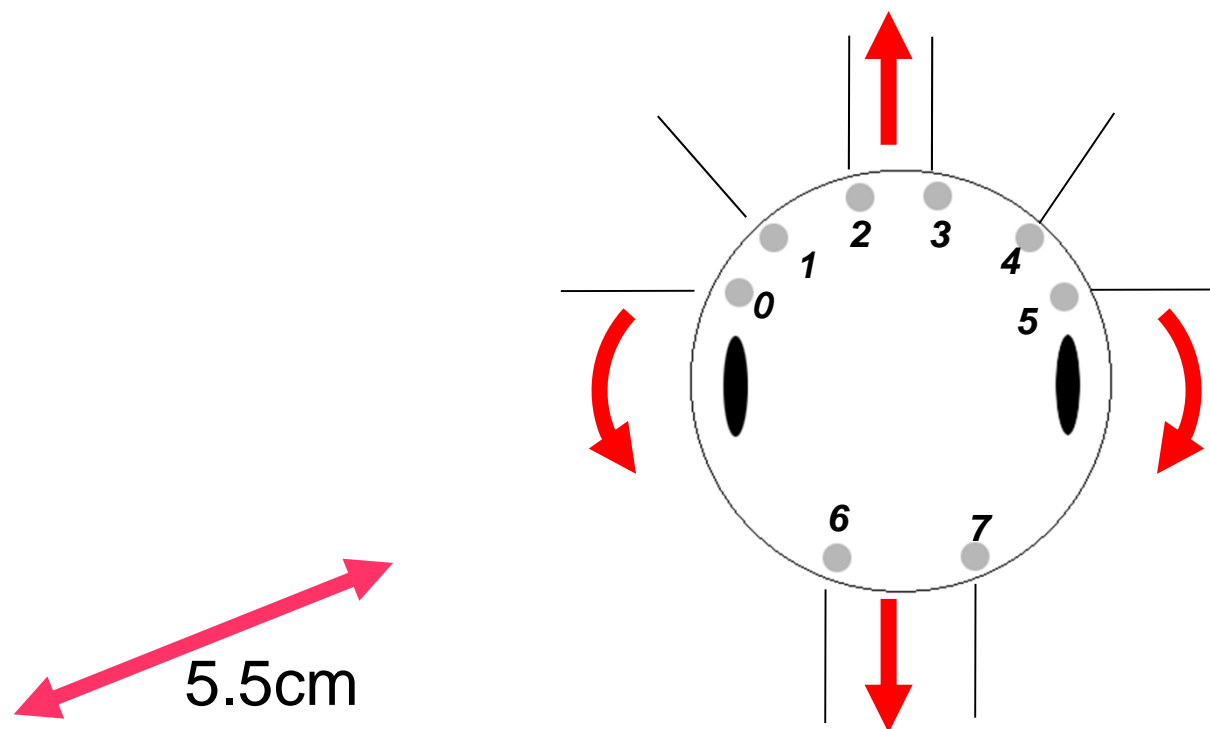
- 強化学習は赤ちゃんの学習と似ている
- 赤ちゃんは親にほめてもらえるような行動をとる



ロボットの障害物回避(1)

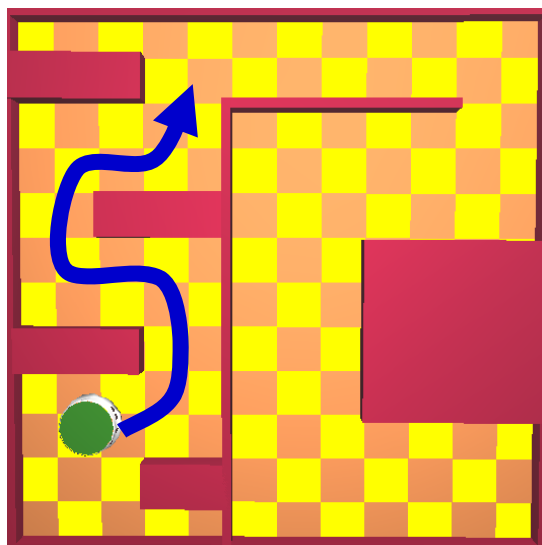
■ Kheperaロボット

- 2つの車輪: 前進, 後進, 左回転, 右回転
- 8個の赤外線センサー: 壁との距離を計測

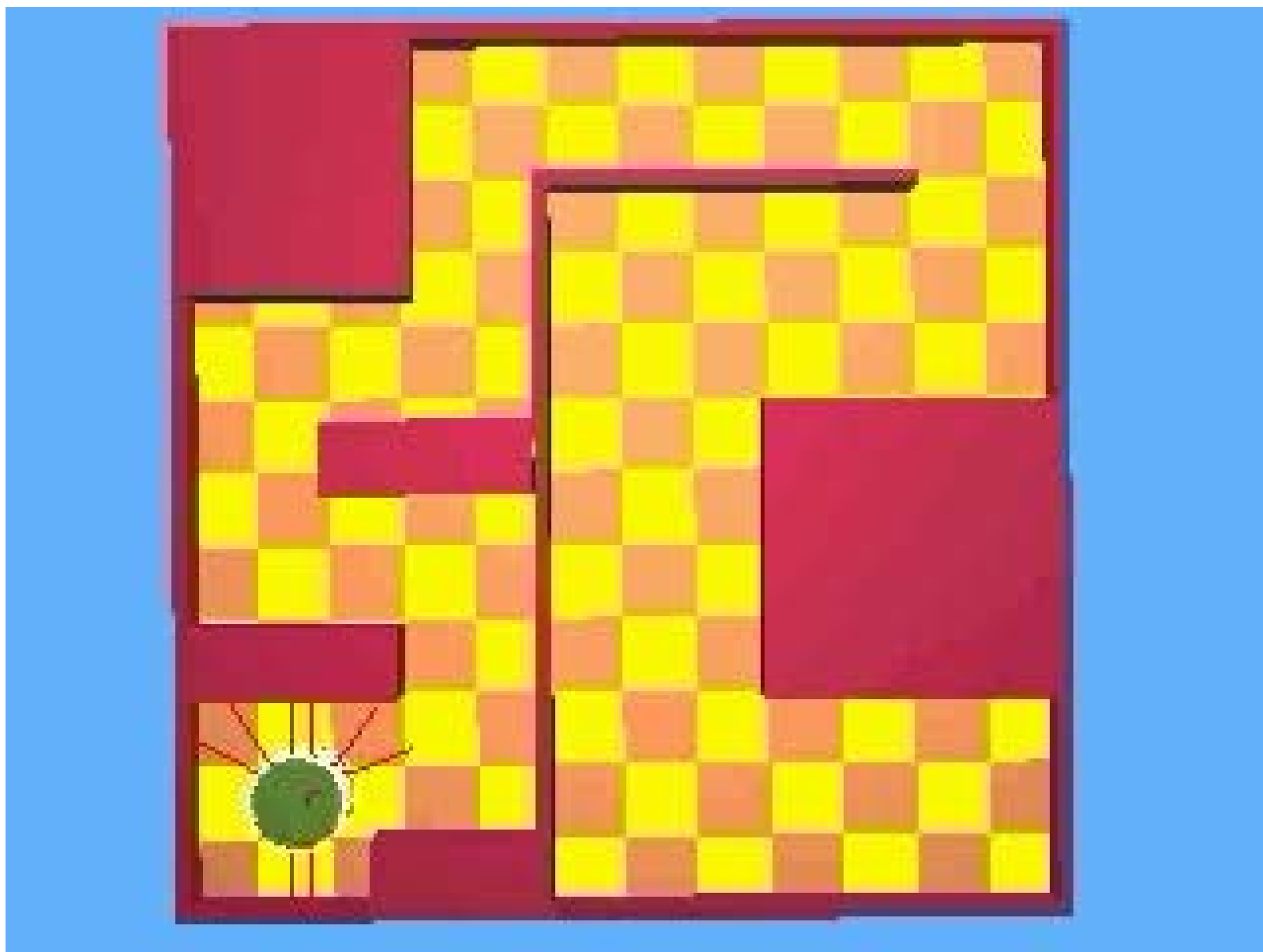


ロボットの障害物回避(2)

- **ゴール**: 障害物を避けて進む
- **報酬**:
 - 壁にぶつかったらマイナスの報酬(罰金)
 - 前に進んだらプラスの報酬
- **どうやって障害物を避けるかは教えない!**



デモ



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＝頭的位置が最高

デモ(1)

■ 学習前

デモ(2)

■ 750回の学習後

デモ(3)

■ 920回の学習後

宿題

- TokyoTech OpenCourseWareのページに確率と統計の基礎に関する追加の資料を公開する:

<http://www.ocw.titech.ac.jp/>

- そこに含まれる宿題を行うこと.
 - 宿題はA4のレポート用紙にまとめること.
 - 学科, 学籍番号, 氏名を忘れず記入すること.
 - 宿題は次の授業の開始時に集める.
 - 遅刻提出は減点するので注意すること.