

き か い が く し ゅ う  
**機械学習**    だい    かい  
**第1回 機械学習とは**

り つ め い か ん だ い が く    じ ょ う ほ う り こ う が く ぶ  
**立命館大学 情報理工学部**

ふ く も り    た か ひ ろ  
**福森 隆寛**

**Beyond Borders**

# 授業の概要 (シラバスより)

□ 機械学習は、アルゴリズムとして明示的に解法が与えられないタスクに対して、そのタスクを遂行するためのモデルを学習データから構築する処理である

□ 本科目では

- データ収集から結果の評価に至るまでの機械学習の実装手順
- 教師あり学習 (識別や回帰など)
- 教師なし学習 (モデル推定やパターンマイニングなど)
- 半教師あり学習
- 強化学習

など機械学習に必要な理論や  
基本的な手法を理解することを目標とする

# 到達目標（シラバスより） と う た つ

## □ 機械学習の一連の手順 い ち れ ん

- データの収集、前処理、評価基準の設定、  
学習、結果の評価  
ま え し ょ り      き じ ゅ ん      せ っ て い

を理解できる

## □ 与えられたタスクに対して あ た

- 対象とするデータの種類  
たいしやう      し ゅ る い
- 最終的に得たい結果  
さいしやうてき      え
- 機械学習手法（教師あり学習、教師なし学習、  
半教師あり学習、強化学習など）

て き せ つ      く      み      あ      わ      せ  
を適切に組み合わせながら処理できる

# 講義スケジュール

## □ 担当教員 1 : 福森 (第1回～第15回)

1	機械学習とは、機械学習の分類
2	機械学習の基本的な手順
3	識別 ( 1 )
4	識別 ( 2 )
5	識別 ( 3 )
6	回帰
7	サポートベクトルマシン
8	ニューラルネットワーク

9	<sup>しんそう</sup> 深層学習
10	アンサンブル学習
11	モデル推定
12	パターンマイニング
13	<sup>けいれつ</sup> 系列データの識別
14	半教師あり学習
15	強化学習

## □ 担当教員 2 : 叶昕辰<sup>せんせい</sup>先生 (第16回の講義を担当)

# 今回の講義内容<sup>ないよう</sup>

- 人工<sup>じんこう</sup>知能<sup>ちのう</sup>・機械学習・深層<sup>しんろう</sup>学習の違い
- 機械学習とは
- 機械学習<sup>ぶんるい</sup>の分類
  - 教師あり学習：識別、回帰
  - 教師なし学習：モデル推定、データマイニング
  - 中間的学習：半教師あり学習、強化学習
- 演習<sup>えんしゅう</sup>問題<sup>もんだい</sup>

# 人工知能・機械学習・深層学習

## □ 人工知能 (Artificial intelligence)

- 本講義では、「<sup>げんざい</sup>現在、<sup>ち て き</sup>人が行っている<sup>はん だ ん</sup>知的な判断を代わりに<sup>か</sup>  
行う<sup>ぎじゅつ</sup>技術<sup>ていぎ</sup>」と定義
- 人と対話<sup>たいわ</sup>を行うアプリやロボット・自動運転<sup>じどううんてん</sup>・病気の診断<sup>びょうき しんだん</sup>の  
補助<sup>ほじょ</sup>・高度な生産システム<sup>こうど せいさん</sup>などの中心<sup>ちゅうしんてき</sup>的技術として位置づけ  
られる

### 人工知能

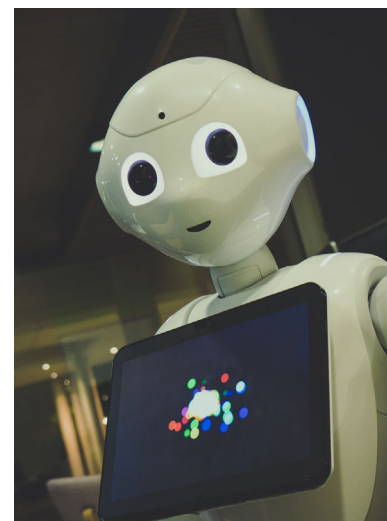
たんさく  
探索  
すいろん  
推論  
ひょうげん  
知識表現

### 機械学習

SVM  
こうばい  
勾配ブース  
ティング

### 深層学習

CNN  
RNN



# 人工知能・機械学習・深層学習

## □ 機械学習 (Machine learning)

■ <sup>かんたん</sup>簡単に<sup>きそくか</sup>規則化できない<sup>ふくざつ</sup>複雑な<sup>たいりょう</sup>大量のデータ (ビッグデータ、Big data) <sup>ゆうよう</sup>から<sup>ちけん</sup>有用な知見を得るための処理

■ 人工知能の技術を実現するための一つの方法 (人工知能 ≠ 機械学習)

- ・ 知的に<sup>ふるまう</sup>振る舞う人工知能システムを作る手段は<sup>つくるしゅだん</sup>機械学習だけではない。<sup>かいはつしゃ</sup>開発者が<sup>さくせい</sup>振る舞いの規則をプログラムとして作成すれば<sup>かのう</sup>実現可能

### 人工知能

探索  
推論  
知識表現

### 機械学習

SVM  
勾配ブース  
ティング

### 深層学習

CNN  
RNN

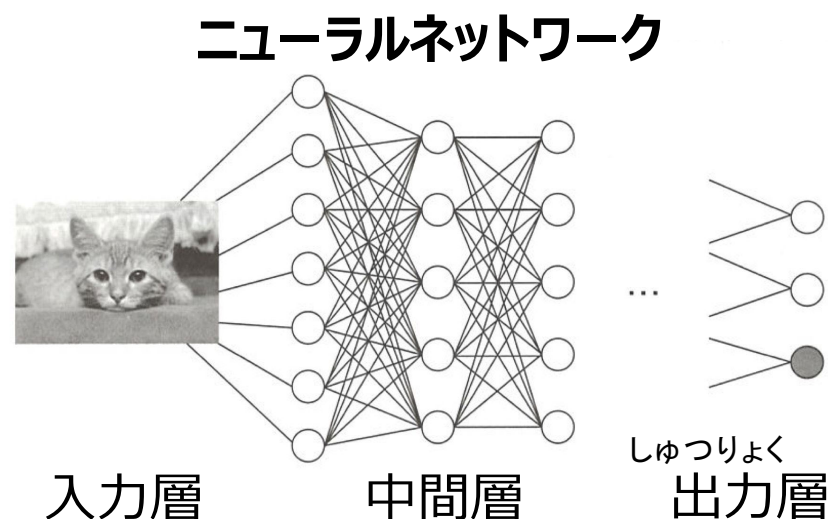
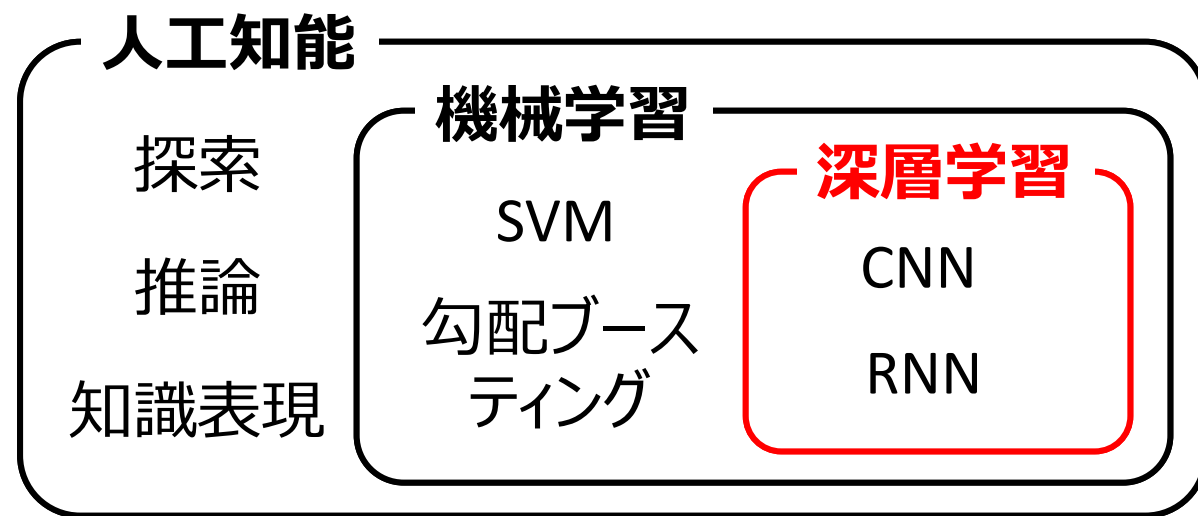
### 深層学習以外の手法も重要

勾配ブースティングなどが<sup>よ</sup>良い<sup>ばあい</sup>場合がある。深層学習より<sup>ていせいのう</sup>低性能でも、<sup>はんてい</sup>判定結果に至るプロセスが<sup>わ</sup>分かりやすい手法や、チューニングが<sup>ようい</sup>容易な手法が<sup>この</sup>好まれる場合もある。

# 人工知能・機械学習・深層学習

## □ 深層学習 (Deep learning)

- 中間層を多くもつニューラルネットワークによって実装
- 機械学習手法の一つで、深い階層構造（多段階の非線形変換）によって、与えられた問題に適したデータの表現法を獲得
- 音声・画像・自然言語など空間的・時間的に局所性をもつ入力対象で、かつ学習データが大量にあることが前提





# 演習問題1-1 (5分間)

□ ビッグデータを用いた機械学習の応用事例を最低  
1つ考えなさい。

例) 翻訳システム  
2ヶ国語で記述された文章データを大量に集めて、そのビッグ  
データから翻訳のルール（規則）を機械学習によって導き出す。  
みちびきだす

# 機械学習とは

## □ 機械学習の基本的な定義

从训练数据中构建一个模型，以执行一项没有给出明确解决方案的任务，作为一种算法。

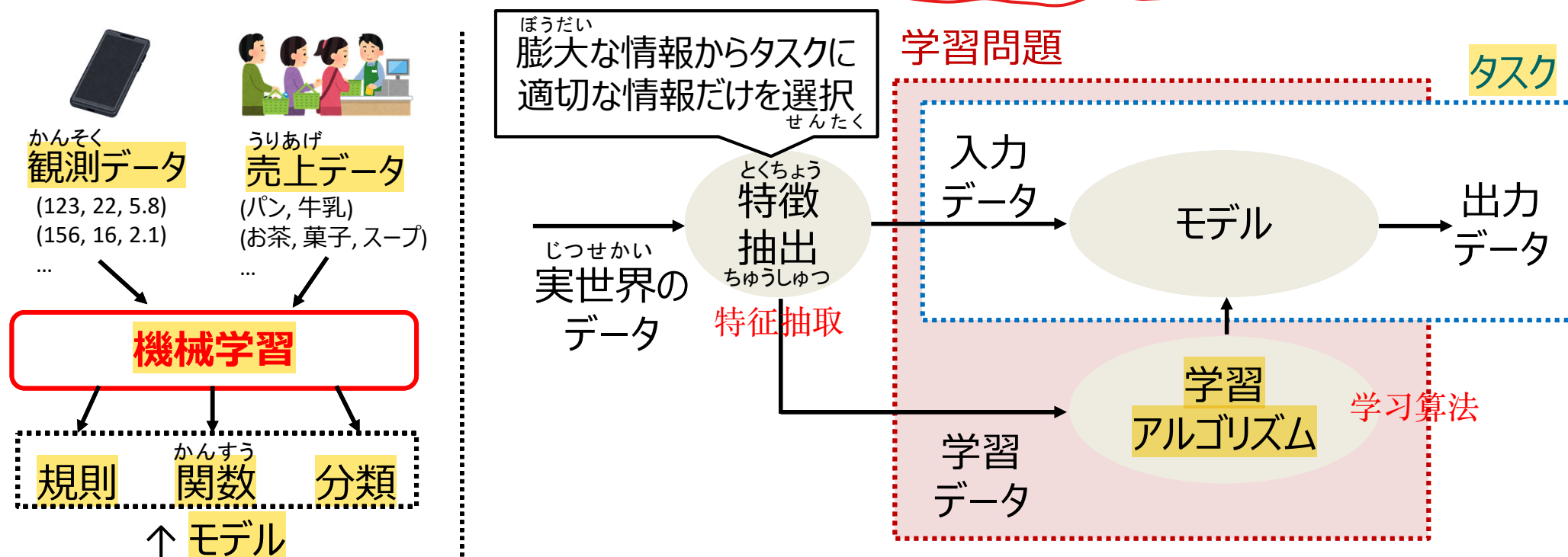
## ■ アルゴリズムとして明示的に解法が与えられていないタスクを

すいこう

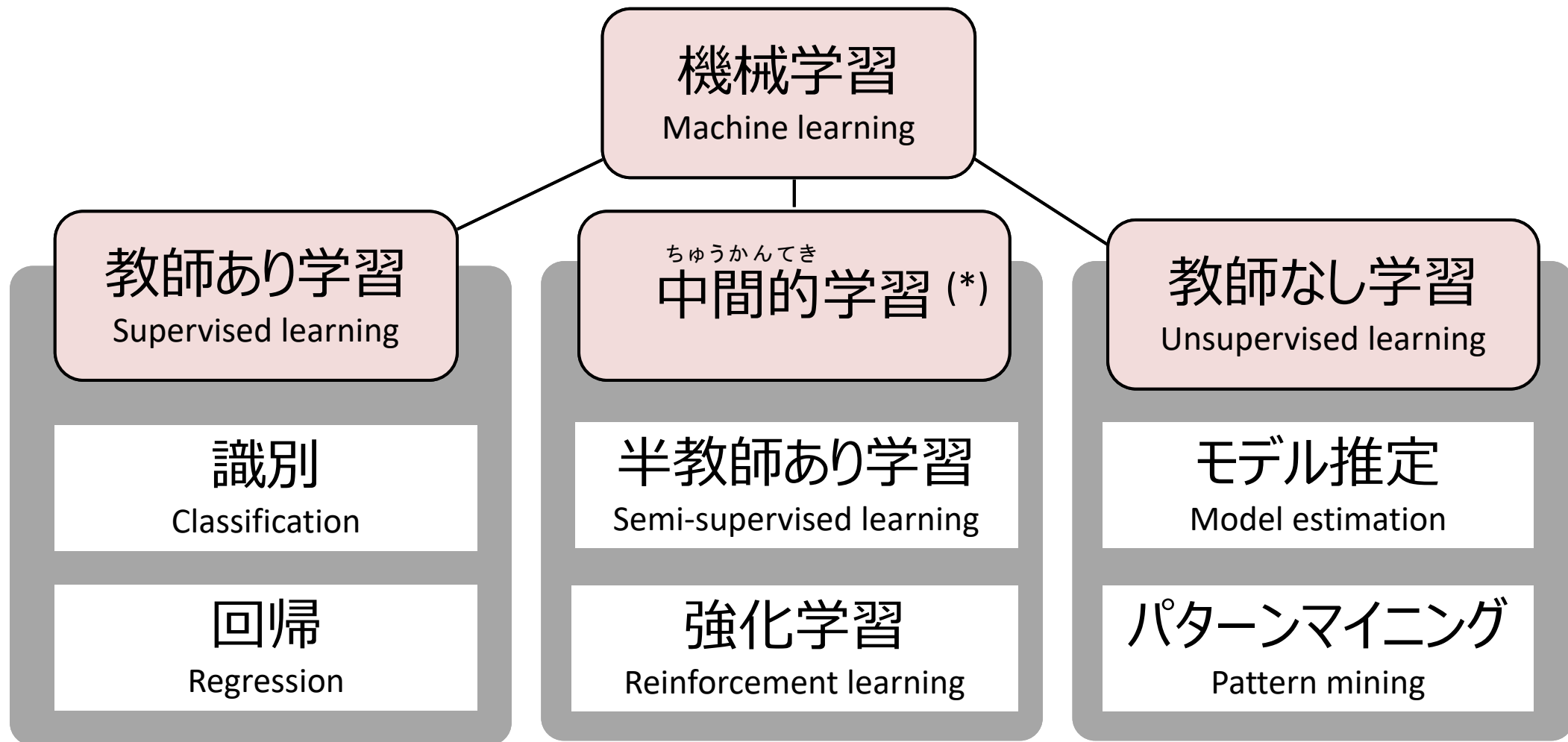
遂行するためのモデルを学習データから構築すること

Task

- ・ **タスク**：機械学習が対象とする問題
- ・ **学習データ**：モデルを構築するために用いるデータ



# 機械学習の分類



(\*) この講義では、「教師あり学習/教師なし学習のいずれにも該当しない手法」を  
中間的学習とします がいう

# 教師あり学習 (Supervised learning)

教師あり学習： 正解の付いた学習データを用いた学習

学習データは、以下のペアで構成

入力データに対応する次元数 $d$ のベクトル  $\{x_i, y_i\}$ ,  $i = 1, 2, \dots, N$   
正解情報 学習データの総数

$$x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T$$

特徴 (feature) と呼び、数値データ (温度や個数など) や  
カテゴリデータ (性別、天気など) が代入される

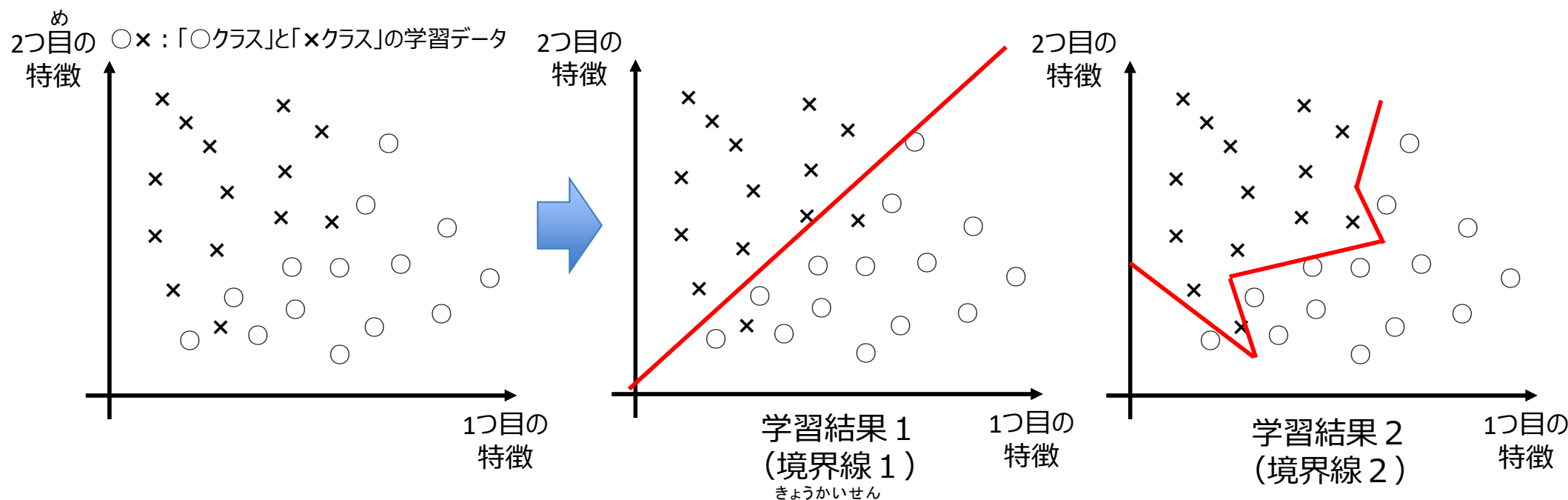
教師あり学習の目的は、入力 $x$ を正解 $y$ に写像する関数 $c$ を推定すること

$c(x): x \rightarrow y$   
この正解が  $\left\{ \begin{array}{l} \text{カテゴリデータの場合: 識別問題} \\ \text{数値データの場合: 回帰問題} \end{array} \right.$

# 識別 (Classification)

## □ 識別問題 将输入分类为预定类别之一的问题

- 入力をあらかじめ<sup>ただ</sup>定められたクラス<sup>もじ</sup>のいずれかに分類する問題<sup>しっぺい う む</sup>
- 応用事例：音声や文字の認識、疾病の有無判定など



入力が2次元数値ベクトルの識別問題  
(○と×のクラスを分ける境界線を考える)

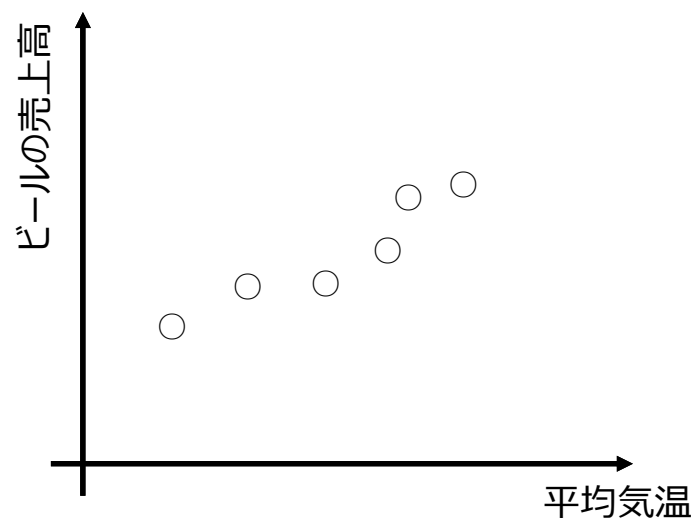
学習データ<sup>かんぜん</sup>を完全に分類するのではなく、未知データ<sup>みち</sup>を<sup>たか</sup>高い  
正解率<sup>せいかい</sup>で分類する境界線を探ることが重要。今回の場合、  
境界線 1 が良いと考えることが多い。

# 回帰 (Regression)

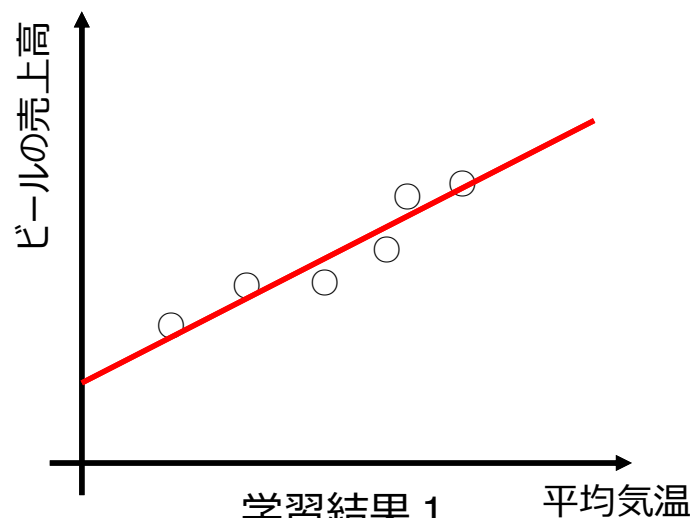
## □ 回帰問題 从输入中找到期望的合理输出值

- 入力から 予想される妥当な出力値 を求める問題
- 応用事例：消費電力の予測、中古車の価格算出など

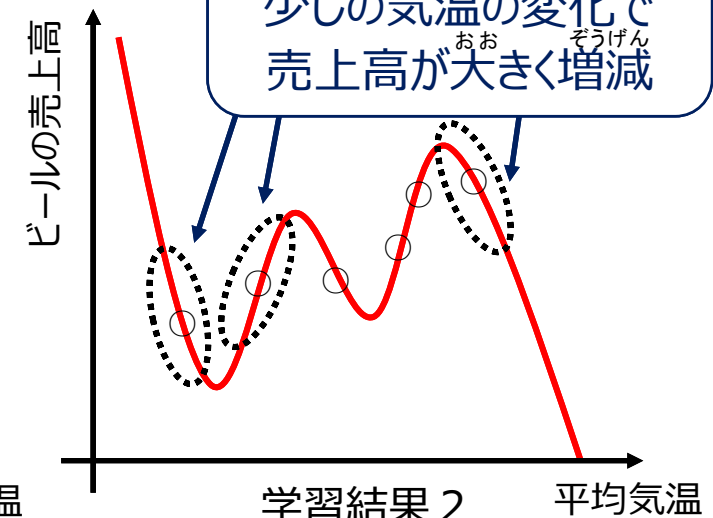
○：学習データ (過去の平均気温とビールの売上高)



平均気温とビールの売上高の回帰問題  
(気温から売上高を予測する関数を考える)



学習結果 1  
(1次式による回帰)



学習結果 2  
(高次式による回帰)

学習データの全ての点を通るのではなく、未知データに対して 妥当な数値を出力する関数を考えることが重要。今回の場合、1次式による回帰が良いと考えることが多い。

# 教師なし学習 (Unsupervised learning)

教師なし学習： 正解が付いていない 学習データを用いた学習

学習データは、以下で構成

$$\{\mathbf{x}_i\}, \quad i = 1, 2, \dots, N$$

入力データに対応する次元数 $d$ のベクトル

学習データの総数

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id})^T \leftarrow \text{特徴：数値データやカテゴリデータ}$$

教師なし学習の目的は、入力データに<sup>ひそ</sup>潜む<sup>せい</sup>規則性を学習すること

① **モデル推定** 将輸入数据整体进行整合的规律

「入力データ<sup>ぜんたい</sup>全体を<sup>しはい</sup>支配する規則性」を学習によって推定する問題

② **パターンマイニング** 在输入数据的子集内或子集之间建立的规律

「入力データの<sup>ぶぶんしゅうごうない</sup>部分集合内、あるいはデータの<sup>かん</sup>部分集合間<sup>なりたつ</sup>に成り立つ規則性」

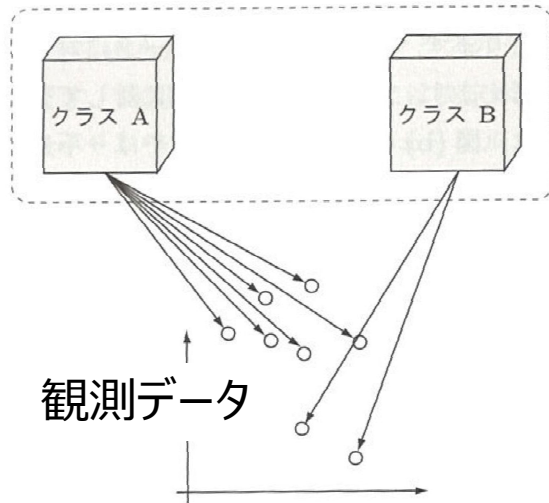
を発見する問題

はっけん

# モデル推定 (Model estimation)

用输入数据的一些共同点来拼凑数据。きょうつう

- 入力データから何らかの共通点をもつデータをまとめること
- 入力データを生じさせたクラスの存在やそのパラメータを推定



モデル推定の考え方

観測されたデータは、何らかのクラスに属していたものが、揺らぎを伴って生成されたものと考える。  
そく  
ゆ ともな



ぎゃく たど でき  
その逆のプロセスを辿ることが出来れば、  
データを生成するためのクラスを推定できる。

观察到的数据被认为是在某类数据的波动下产生的。

## クラスタリング (Clustering)

入力データ集合から適切なまとまりを作ることによってクラスを推定する手法

## 密度推定 (Density estimation)

クラスはデータを生成する関数を持っていると仮定して、その関数のパラメータを入力データから推定する手法

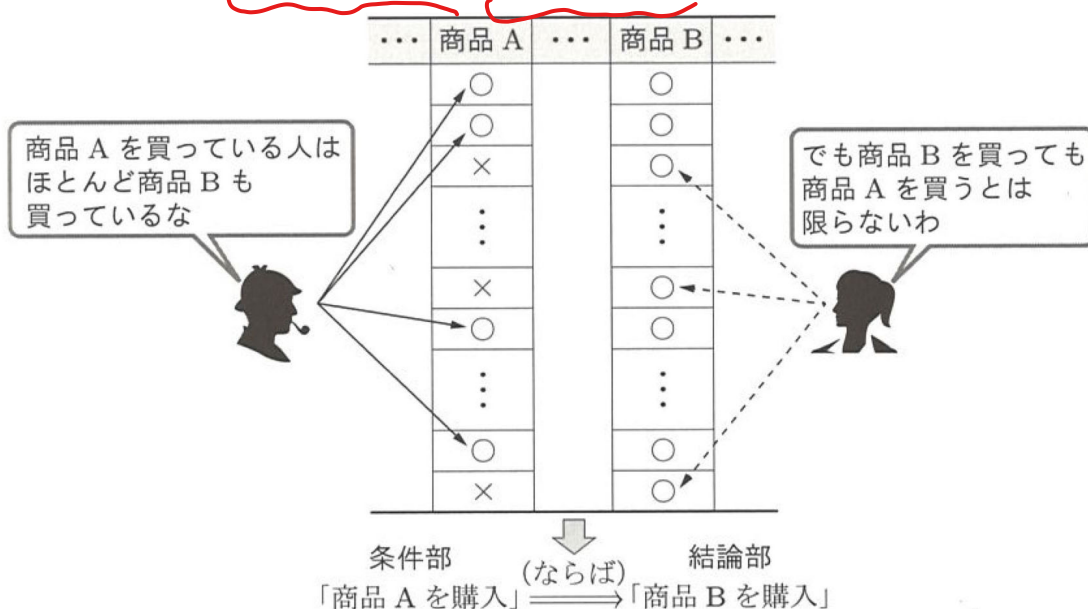


# パターンマイニング (Pattern mining)

□ 大量のデータから、<sup>なんど しゅつげん</sup> 何度も出現するパターンや、その  
<sup>もと</sup> パターンに基づいた規則を発見 (= <sup>はっくつ</sup> 発掘 : マイニング) すること

<sup>て き</sup> <sup>けいさん</sup> パターンマイニングの敵は膨大な計算量

<sup>じょうけんぶ</sup> 条件部 <sup>けつろん</sup> 結論部から構成



売上げデータに対するパターンマイニングの考え方  
(例：店で同時に購入される商品の組み合わせを発見)  
<sup>どうじ</sup> <sup>こうにゆう</sup> <sup>しょうひん</sup>

<sup>さず</sup> 左図の事例では、発見された  
条件部・結論部の要素数が **1つ**  
<sup>ようそすう</sup>

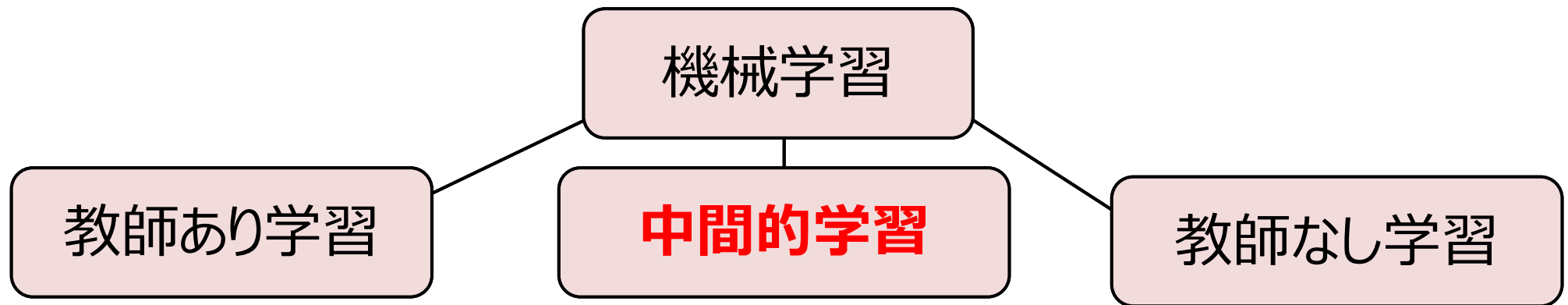
↓  
全ての売上げデータに対して、その  
出現頻度を計算することは容易  
<sup>しゅつげんひんど</sup>

<sup>いっぱんてき</sup> 一般的には、条件部・結論部の  
いずれも **要素の集合**  
↓

あらゆる組み合わせに対して、マイ  
ニングの対象となるデータ集合から  
出現数を数え上げなければならない  
<sup>かぞえあげ</sup>

# 中間的学習

本講義では、中間的学習を「教師あり学習・教師なし学習のいずれにも該当しない手法」と定義



## □ 中間的学習の例

### ■ 半教師あり学習

- 正解付きと正解なしの学習データが混在している

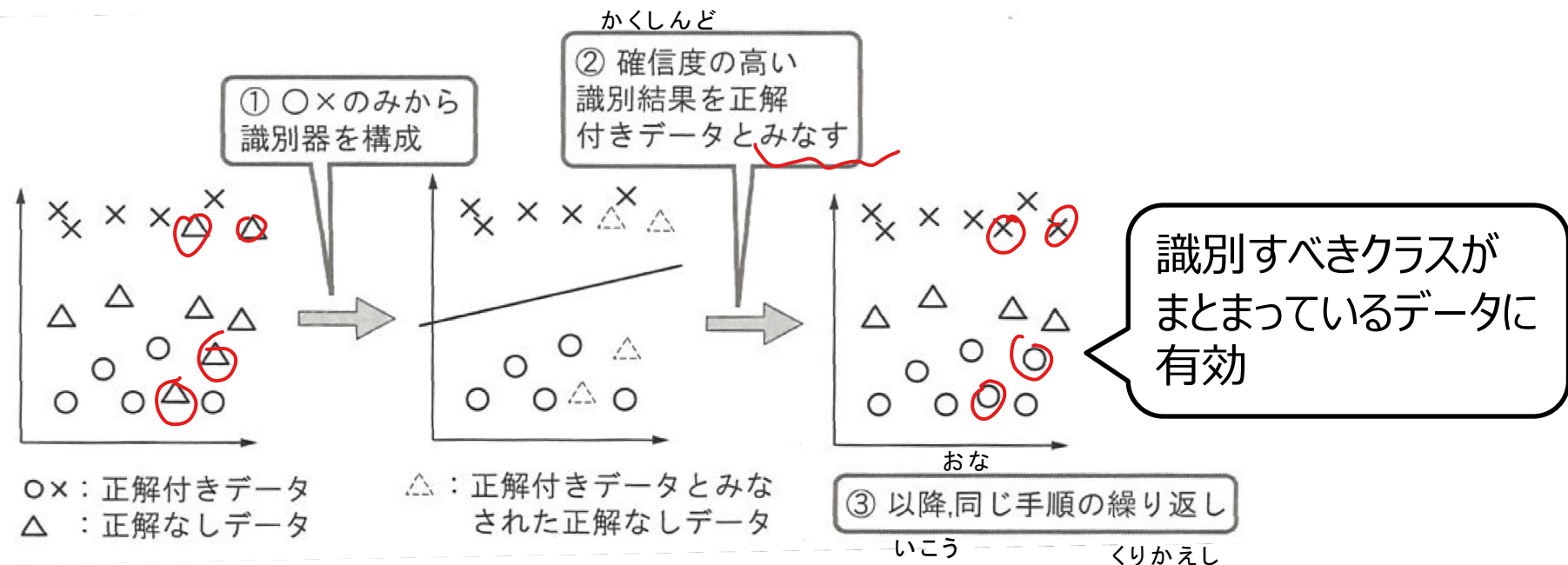
### ■ 強化学習

- 与えられる正解が間接的で、教師あり/教師なし学習の中間的な状況

# 半教師あり学習 (Semi-supervised learning)

正解付きデータから得られた識別器の性能を、正解なしデータを使って向上させる手法 使用无标签数据来提高由有标签数据训练得到的分类器的性能

正解付きデータの作成のために必要な人手・費用・時間に限界があるが、正解なしデータであれば、容易かつ大量に入手できるような状況で有効な手法



半教師あり学習の考え方  
(主に識別問題に対して用いられる)

# 強化学習 (Reinforcement learning)

一連の行動を通じて、最も多くの報酬を得られるように、  
各状態における最適な出力を獲得させる学習手法

報酬を教師信号とすれば、教師あり/なしの中間的な設定と考えることができる。  
個々の決定に正解は与えられず、決定の連続に対して、あとから間接的な教師信号が与えられる。

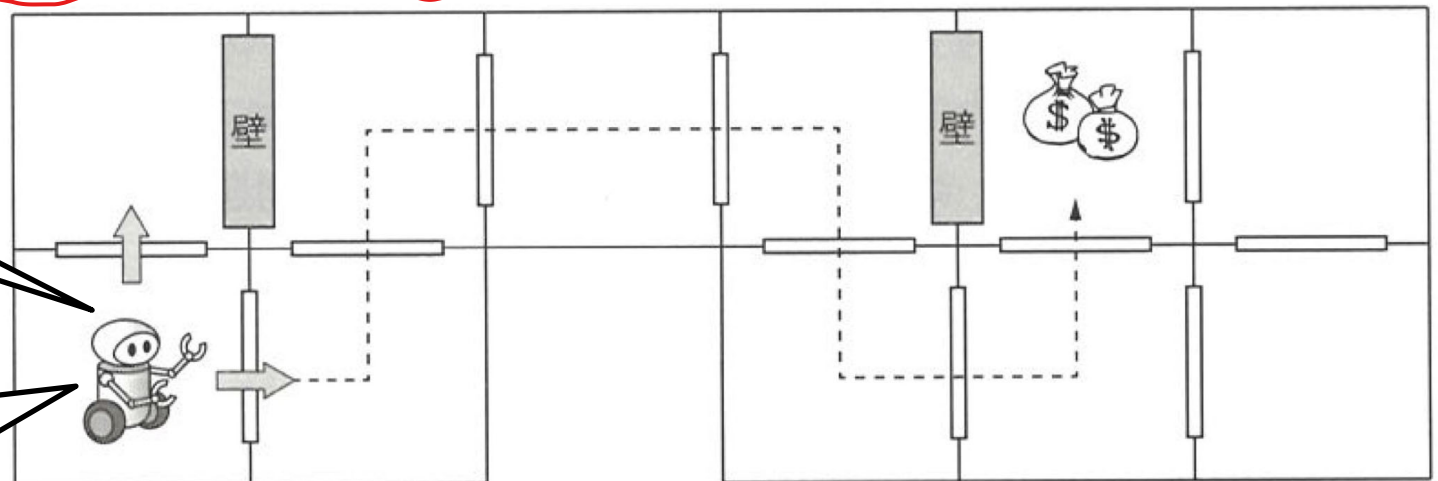
ロボットに試行錯誤を繰り返させながら、ゴールに到着したときだけ報酬を与える

## 入力

ロボットのセンサ情報  
(ロボットの位置情報)

## 出力

ロボットの移動コマンド  
(上下左右に進む)



強化学習の問題例 (迷路を抜けるロボットを学習)

めいろ

# 演習問題1-2（10分間）

□ 以下の学習手法について応用事例を1つずつ考えなさい。

- 教師あり学習（回帰）
- 教師なし学習（モデル推定）
- 半教師あり学習
- 強化学習

※ 講義スライドに記載されていない応用事例を考えてください。