

1. はじめに

内容

1.1 人工知能・機械学習・深層学習

何が違うか、何ができるか

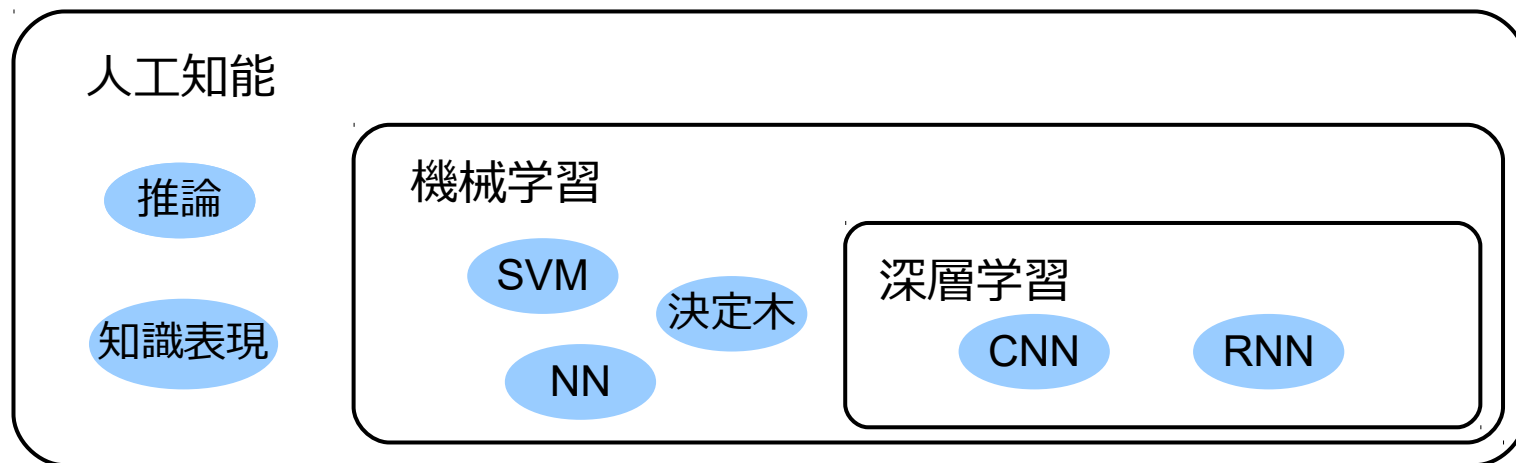
1.2 機械学習とは何か

機械学習の全体像

1.3 機械学習の分類

教師あり学習、教師なし学習、中間的手法

1.1 人工知能・機械学習・深層学習

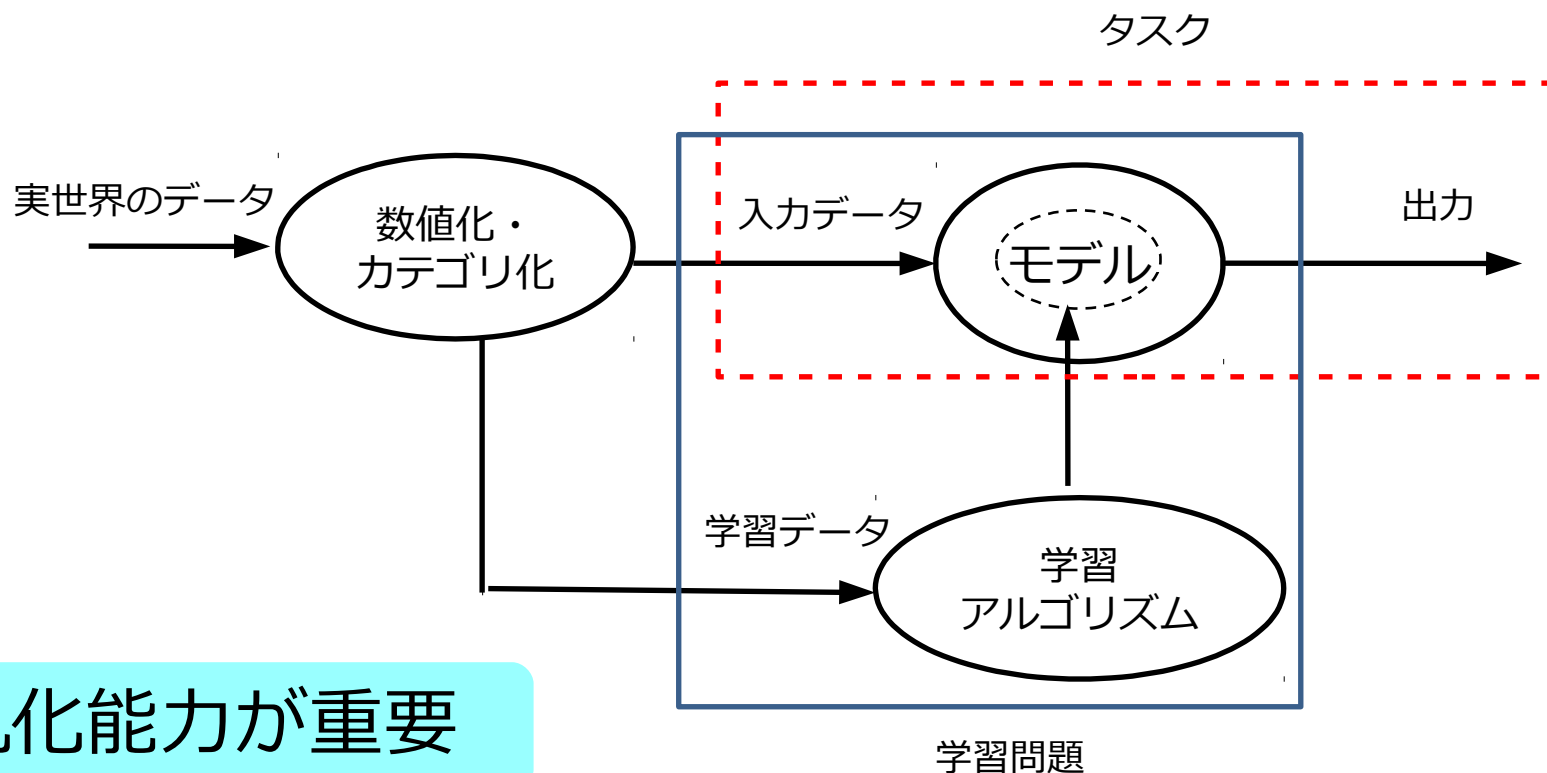


- 人工知能とは
 - 現在、人が行っている知的な判断を代わりに行う技術
 - 技術が普及すると人工知能とはみなされなくなる
 - 例) 文字認識
 - 探索・知識表現・推論・機械学習などを含む

1.1 人工知能・機械学習・深層学習

- 機械学習とは

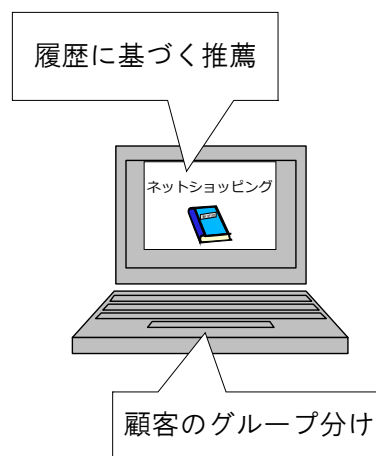
- 機械学習は、適切に**タスク**を遂行する適切な**モデル**を、適切な**特徴**から構築すること [Flach 2012]



1.1 人工知能・機械学習・深層学習

- 機械学習が流行した理由
 - ネットワーク、センサー等の発達によってビッグデータが得られた
 - 計算機の高速化でビッグデータが処理可能になった
- ビッグデータは何に使えるか

- 有用な知見の獲得
- 省力化
- 将来の予測



多様な趣味・嗜好に対応



安心・安全を進化

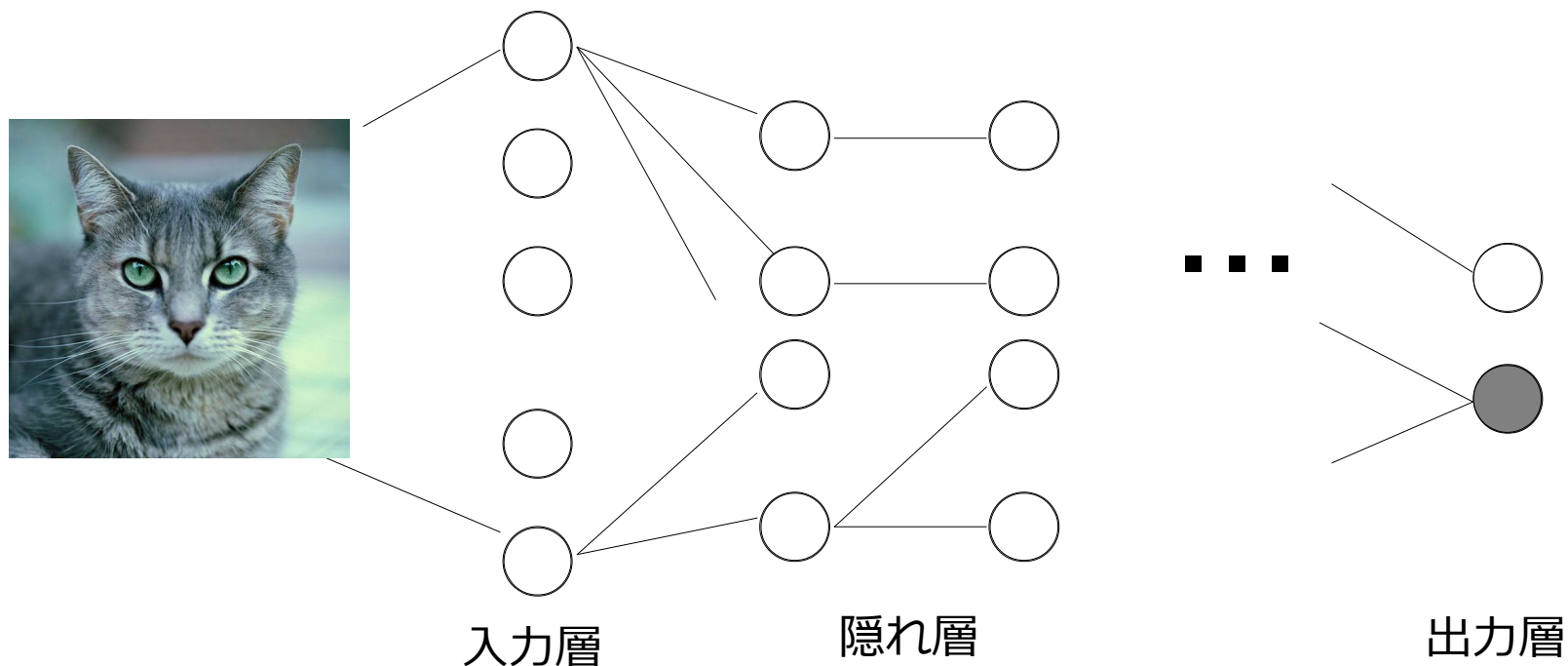


経験や勘を超越

1.1 人工知能・機械学習・深層学習

- 深層学習とは

- 多層に非線形変換を重ねる手法による機械学習
 - 一般的には隠れ層を多くもつニューラルネットワーク
 - 特徴抽出処理も学習対象とすることができる点が特長
 - 学習には大量のデータが必要



1.2 機械学習とは何か

- 機械学習の位置づけ



観測データ

(134.1, 34.6, 12.9)

(135.5, 30.1, 43.0)

...

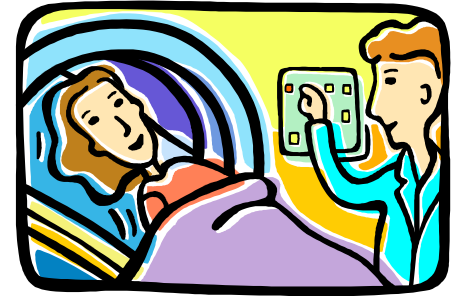


売り上げデータ

(パン、ハム)

(パン、牛乳、バター)

...



診療データ

(男, 28, 178, 75, yes)

(女, 68, 165, 44, no)

...

機械学習

規則

関数

分類

1.2 機械学習とは何か

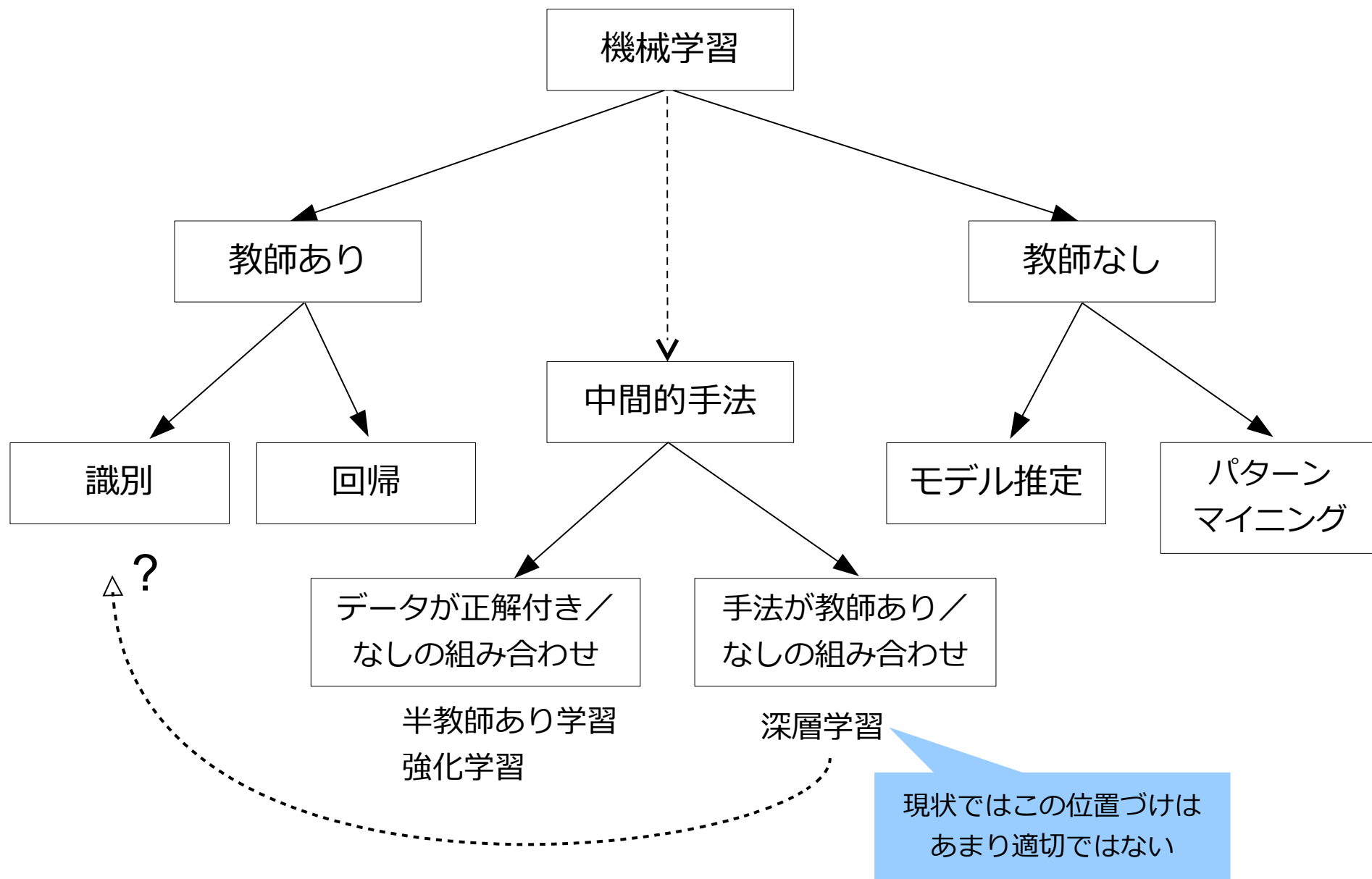
- ビッグデータ活用の古典的事例
 - 紙おむつとビール



1.2 機械学習とは何か

- ビッグデータ活用の近年の事例
 - IBM 社 Watson
 - 2011 年 米クイズ番組 Jeopardy! で人間のチャンピオンをやぶり優勝
 - 2016 年 2000 万件以上の癌に関する論文を学習し、医師が診断できなかった特殊な白血病の発症を見抜く
 - Google 社
 - 2016 年頃、音声認識・機械翻訳の性能が飛躍的に向上
 - 240 万 km を越える自動運転車の公道実走試験実施
 - 2017 年 AlphaGo が世界トップ棋士に三戦全勝

1.3 機械学習の分類



1.3.1 教師あり学習

- 教師あり学習のデータ

- 特徴ベクトル \mathbf{x} と正解情報 y のペア

$$\{(\mathbf{x}_i, y_i)\}, \quad i = 1 \dots N$$

- 特徴ベクトルは次元数 d の固定長ベクトル

$$\mathbf{x}_i = (x_{i1}, \dots, x_{id})^T$$

- 特徴ベクトルの各要素は数値データまたはカテゴリカルデータ

- カテゴリカルデータの例：性別、職業、天候、 etc.

- 正解情報の型によって問題が分かれる

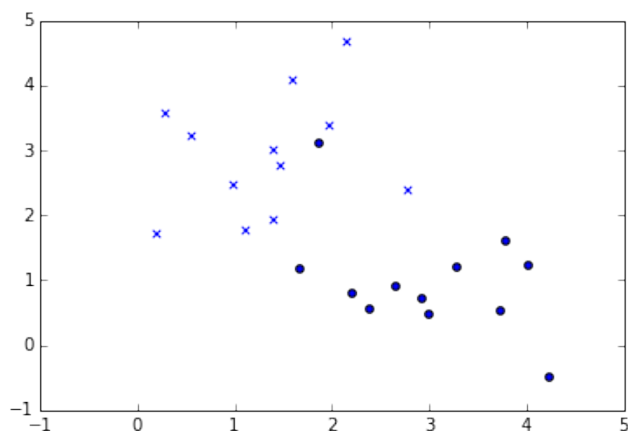
- カテゴリカルデータ：識別

- 数値データ：回帰

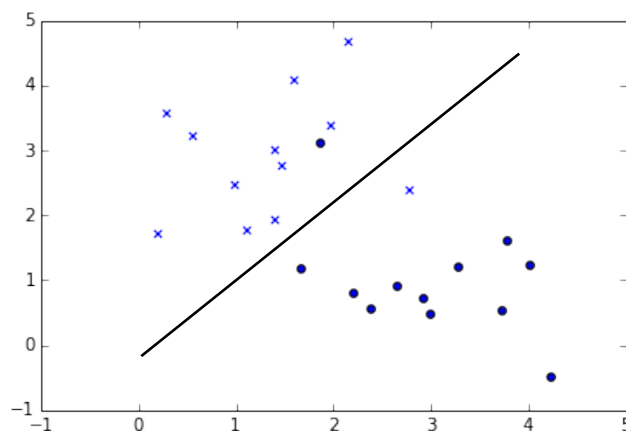
1.3.1 教師あり学習

- 識別

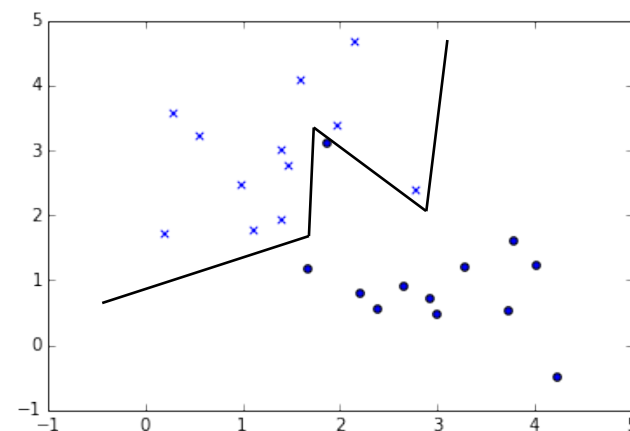
- 正解情報がカテゴリカルデータ
- 汎化誤差が最小となるような特徴空間上の識別面を求める



(a) 入力が2次元数値ベクトルの識別問題



(b) 学習結果 1

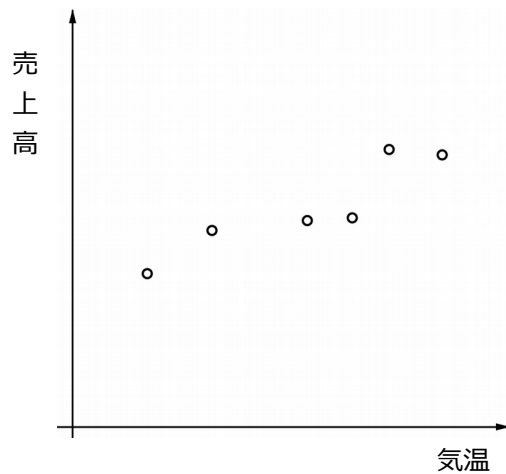


(c) 学習結果 2

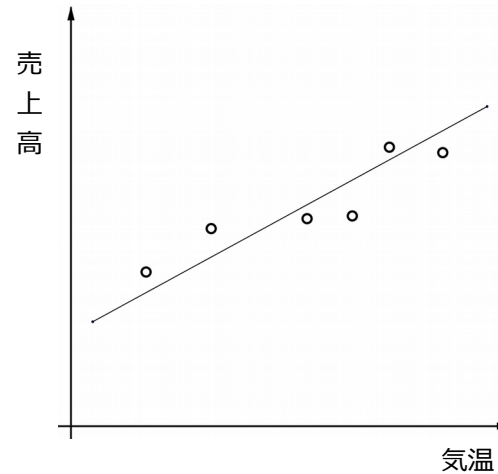
一般化という視点でどちらが適しているか

1.3.1 教師あり学習

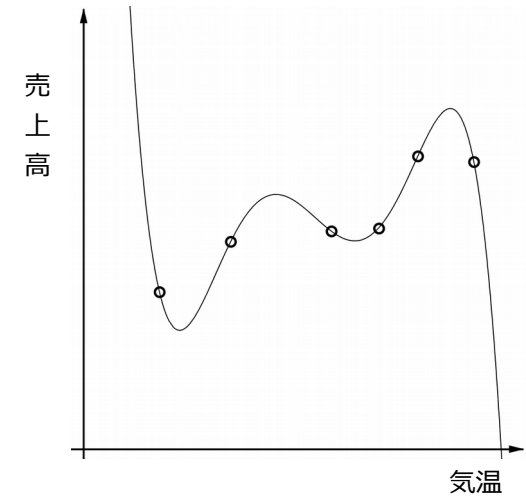
- 回帰
 - 正解情報が数値データ
 - 汎化誤差が最小となるような近似関数を求める



(a) 夏の平均気温とビールの売上高の関係



(b) 1 次式による回帰



(c) 高次の式による回帰

一般化という視点でどちらが適しているか

1.3.2 教師なし学習

- 教師なし学習のデータ

- 特徴ベクトル \mathbf{x} のみ

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

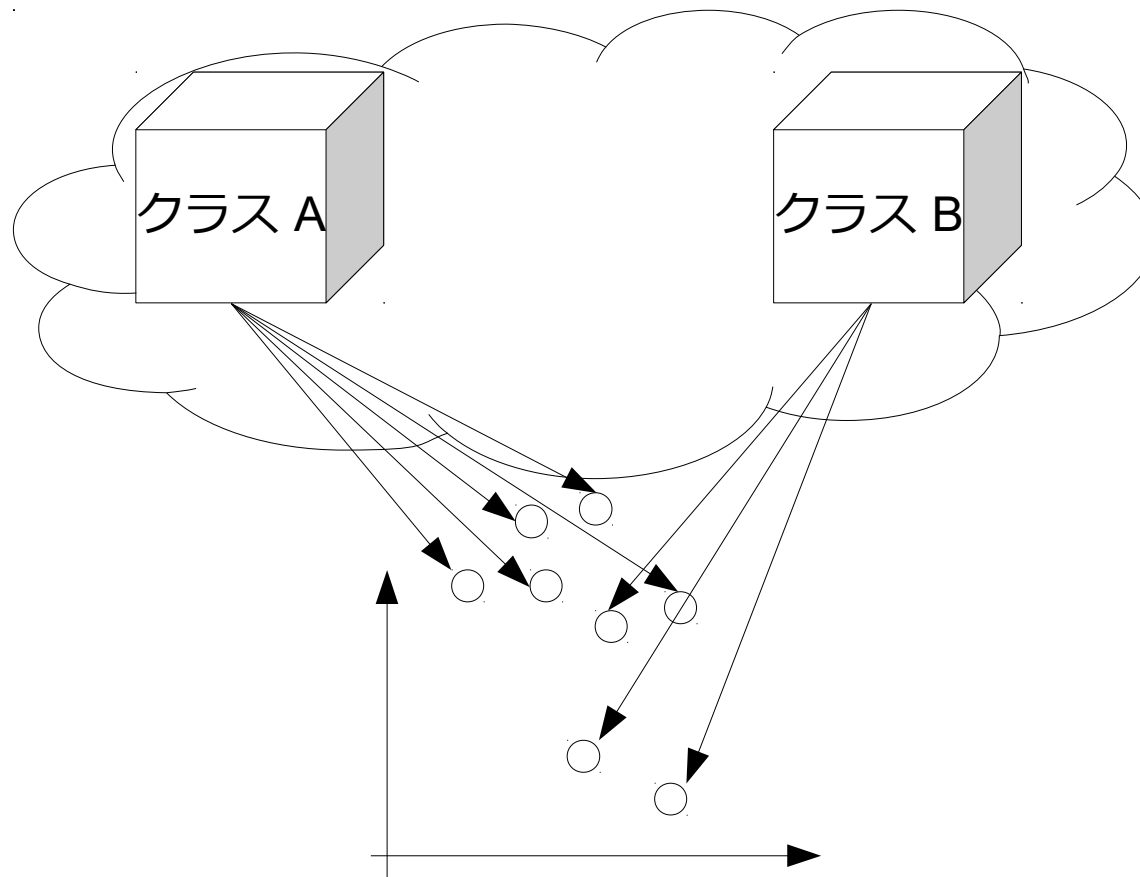
- 特徴ベクトルは次元数 d の固定長ベクトル

$$\mathbf{x}_i = (x_{i1}, \dots, x_{id})^T$$

- 基本的にデータに潜む規則性を学習
- 規則がカバーする範囲によって問題が分かれる
 - データ全体をカバー：モデル推定
 - 頻出する傾向を発見：パターンマイニング

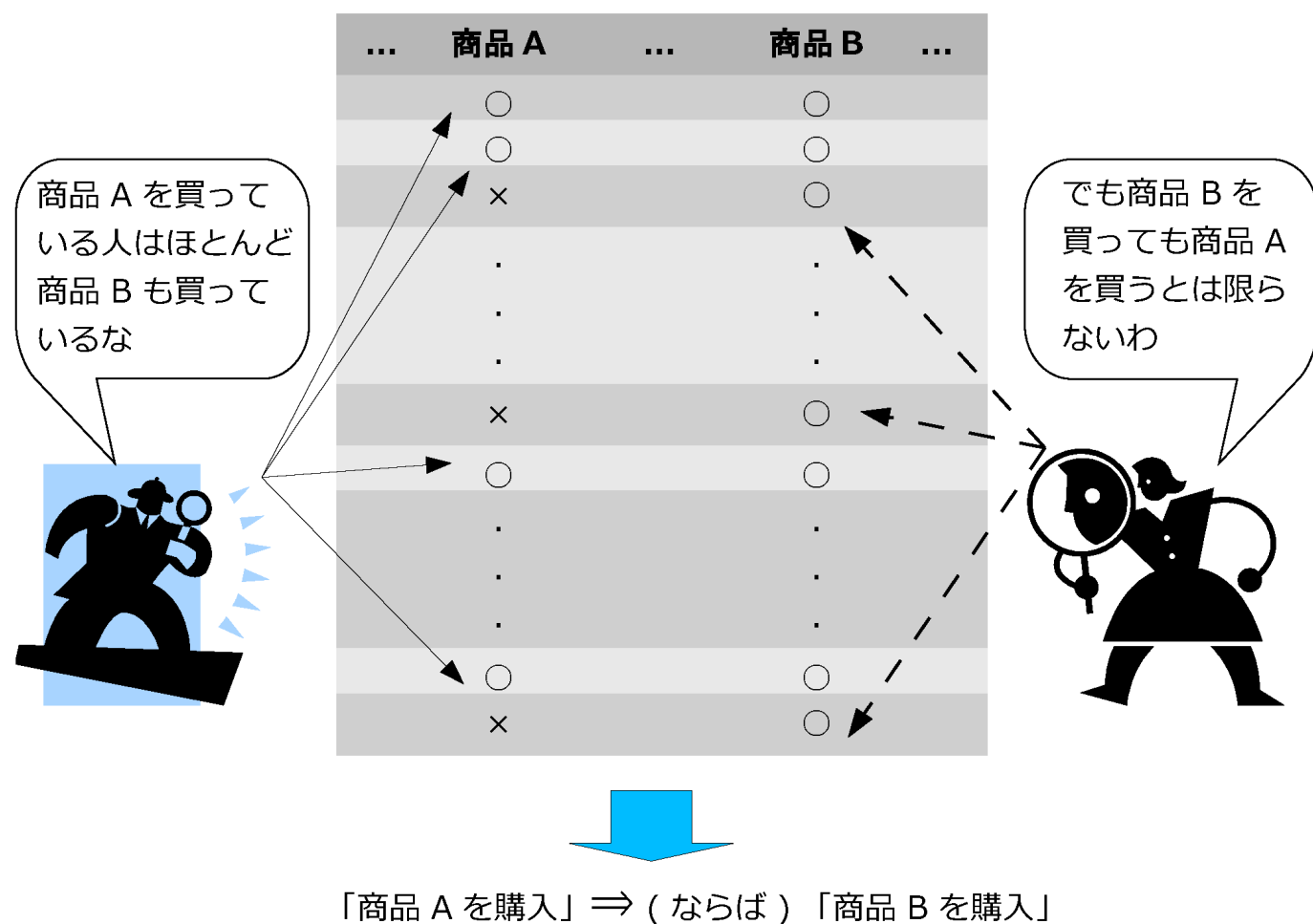
1.3.2 教師なし学習

- モデル推定
 - データを生じさせたクラスを推定
 - 特徴ベクトルは主として数値データ



1.3.2 教師なし学習

- パターンマイニング
 - 頻出項目や隠れた規則性を発掘
 - 特徴ベクトルは主としてカテゴリカルデータ

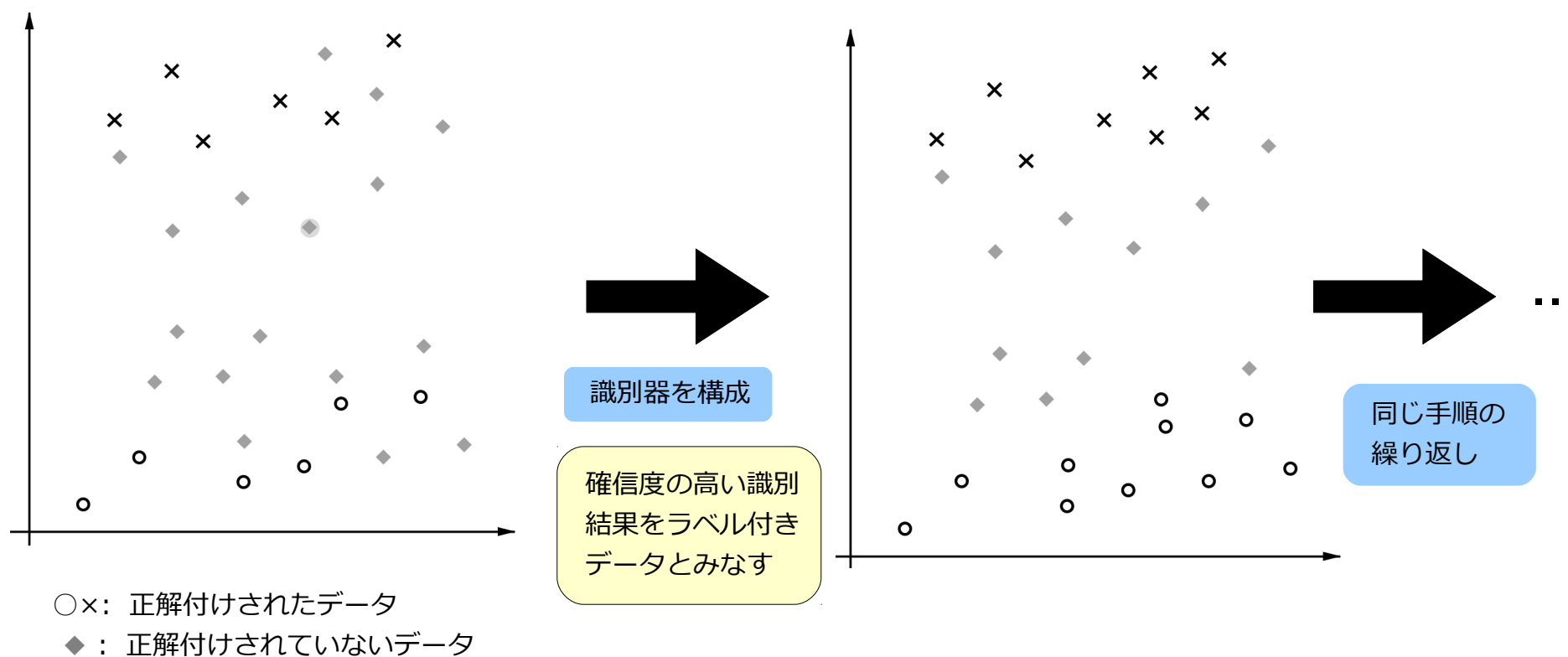


1.3.3 中間的手法

- データが正解付き／なしの組み合わせ
 - 半教師あり学習に適した状況
 - 正解付きの少量のデータ
 - 正解なしの大量のデータ
 - 強化学習
 - 正解情報が、ときどき報酬という形式で与えられる
- 手法が教師あり／なしの組み合わせ
 - 深層学習における事前学習
 - 深層学習における転移学習

1.3.3 中間的手法

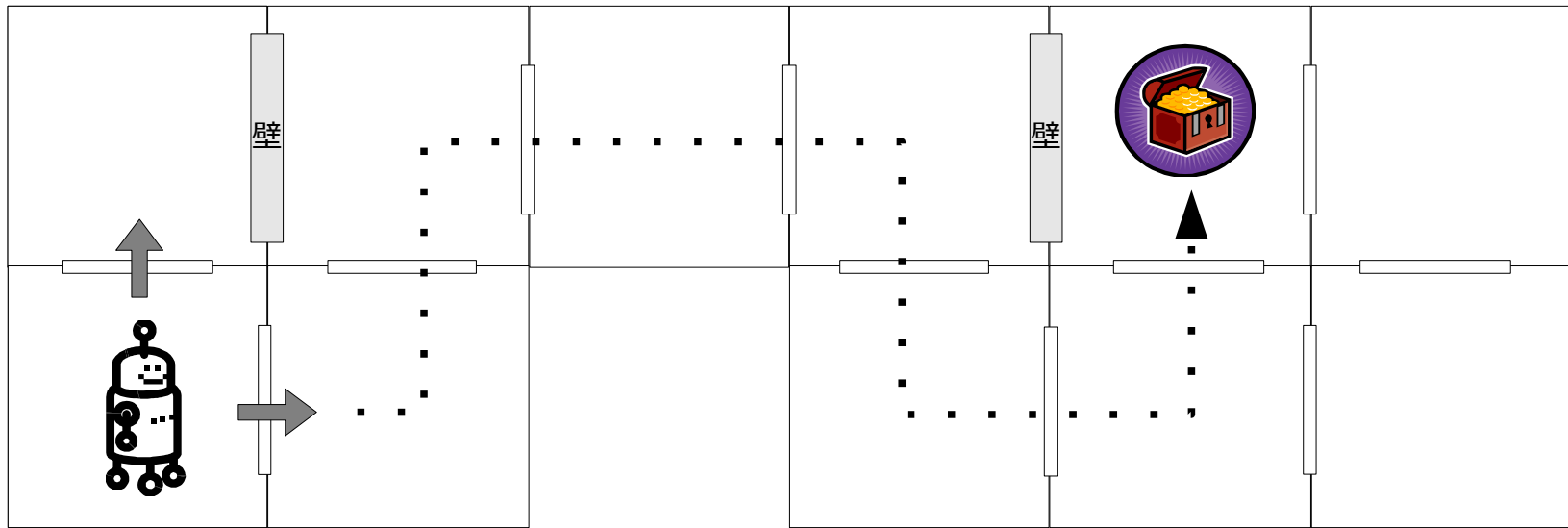
- 半教師あり学習
 - 繰り返しによる学習データの増加



1.3.3 中間的手法

- 強化学習

- 教師信号が、間接的に、ときどき、確率的に与えられる



1.3.3 中間的手法

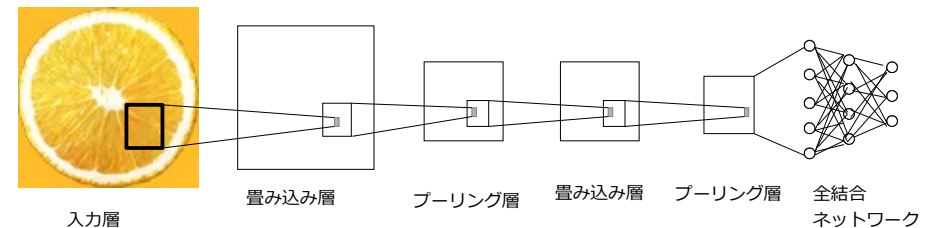
- 深層学習

- 多層に非線形変換を重ねる手法により、特徴抽出を行う表現学習が可能になった

- あるタスクにおける表現学習結果をデータが少ない別のタスクに転用 → 転移学習

- 対象とする問題に応じてさまざまな構造が考案された

- 畳み込みネットワーク



- リカレントネットワーク

