

1. はじめに

内容

1.1 ビッグデータの時代

近年のビッグデータ・機械学習の話題

1.2 機械学習とは何か

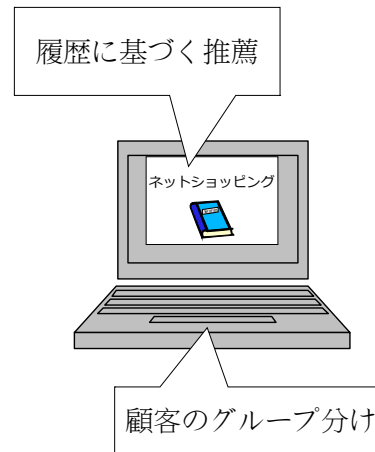
機械学習の全体像

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 ビッグデータの時代

- ビッグデータとは
 - ネットワーク、センサー等の発達によって収集されたデータ
 - 大量・多様・スピードが特徴
- 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



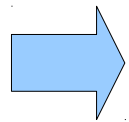
安心・安全を進化



経験や勘を超越

1.1 ビッグデータの時代

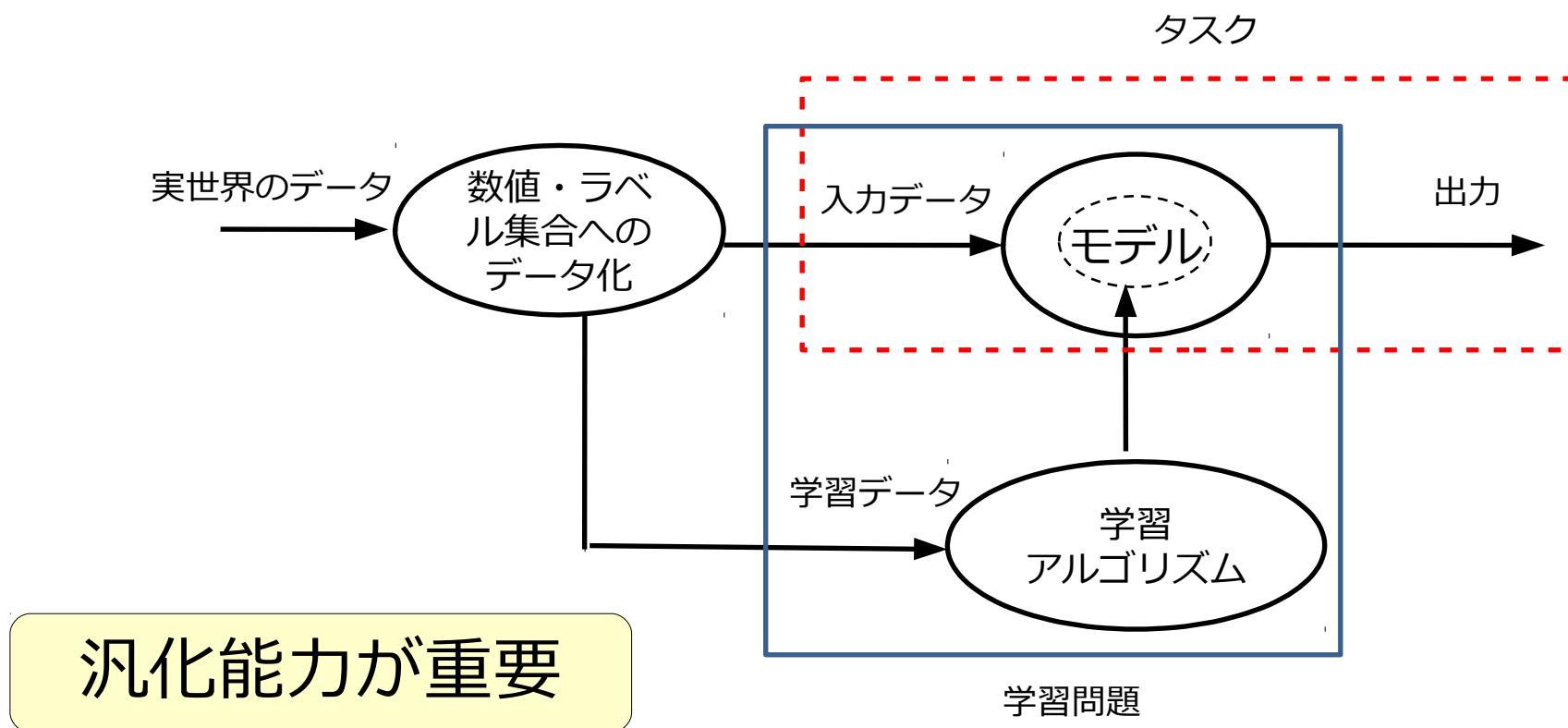
- ビッグデータ処理の問題点
 - データ量が膨大なので人手による情報抽出は不可能
 - 矛盾・曖昧性・ノイズを含むデータを処理するプログラムを記述するのは難しい



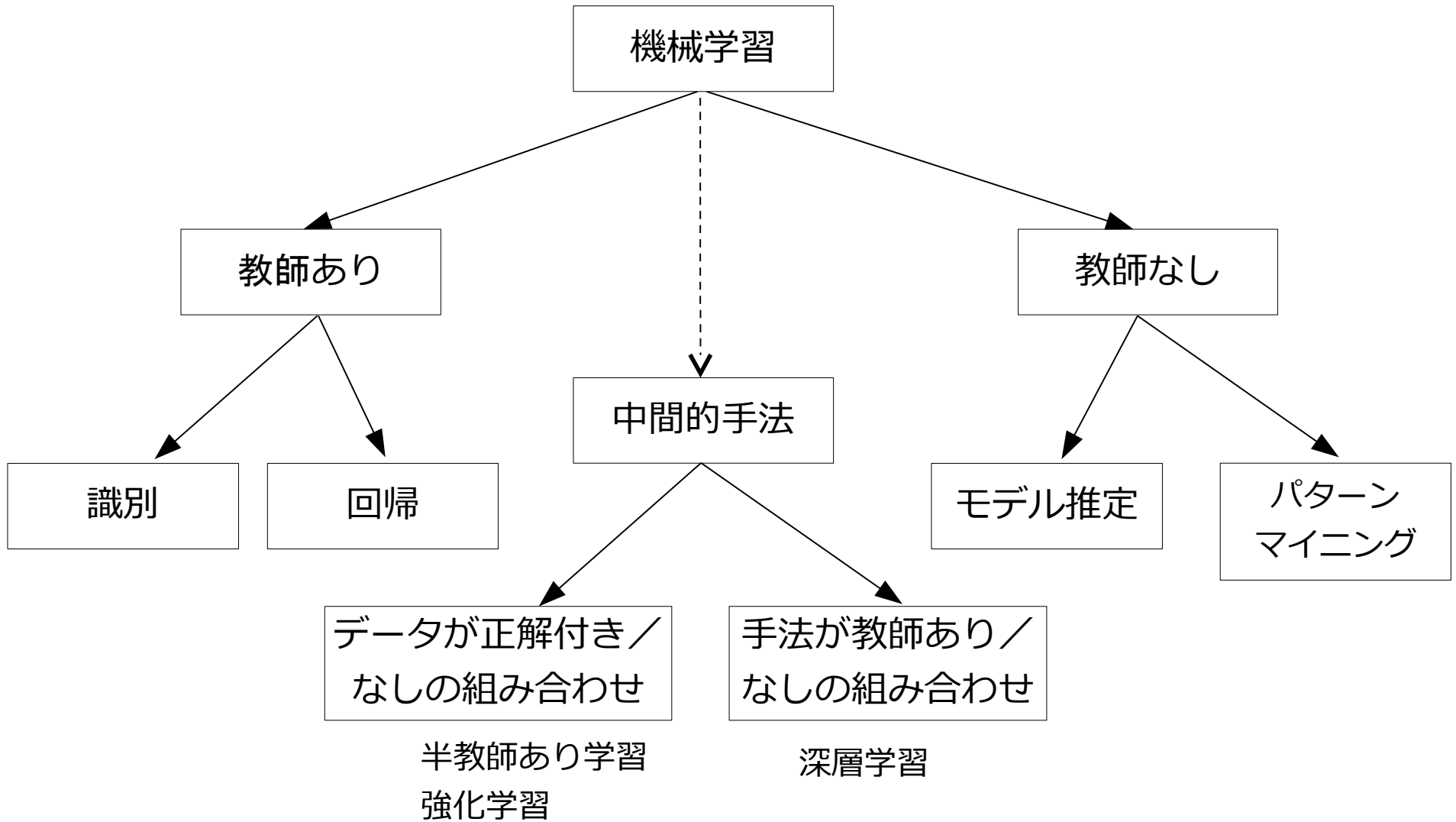
コンピュータによる**機械学習**が有望

1.2 機械学習とは何か

- 機械学習の定義 [Flach 2012]
 - 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



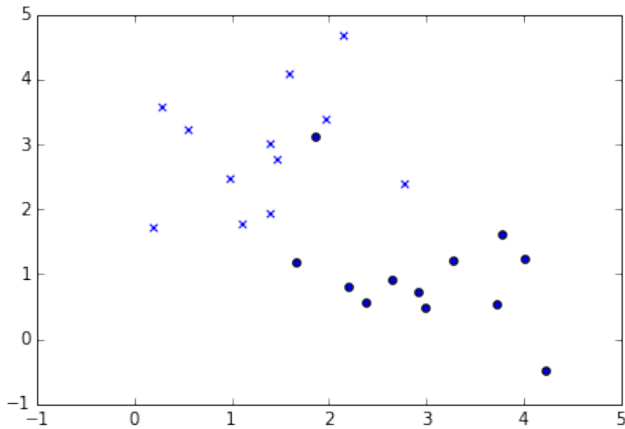
1.3 機械学習の分類



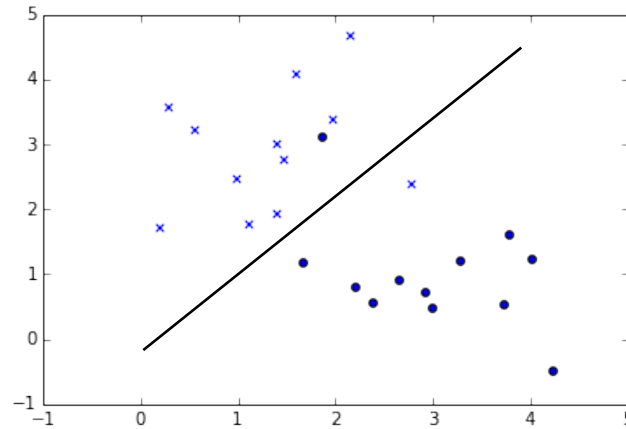
1.3.1 教師あり学習

- 識別

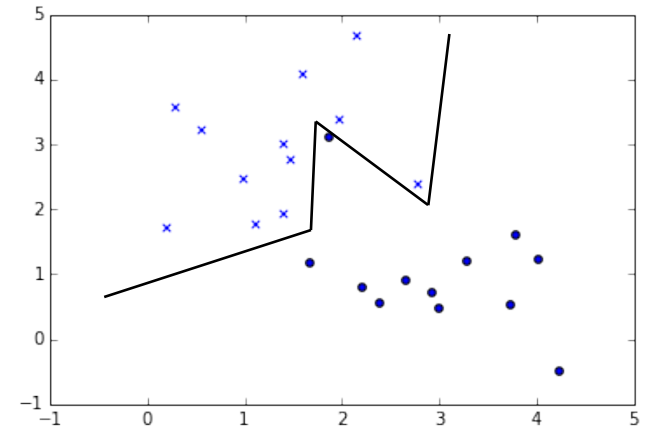
- 学習データに対するエラーが最小となるような特徴空間上の分離面を求める



(a) 入力が2次元数値ベクトルの識別問題



(b) 学習結果 1



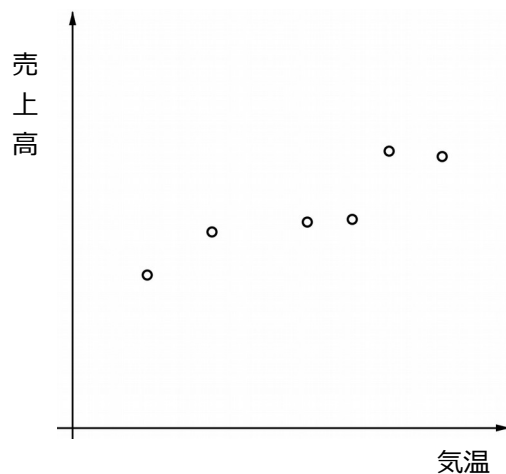
(c) 学習結果 2

一般化という視点でどちらが適しているか

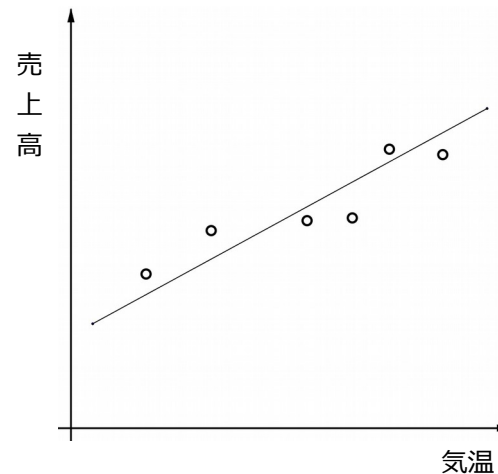
1.3.1 教師あり学習

- 回帰

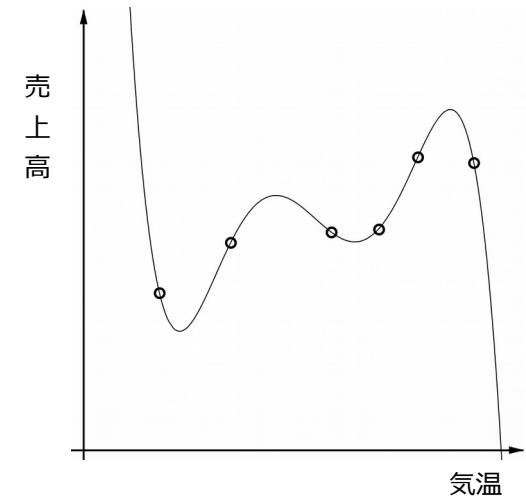
- 学習データに対するエラーが最小となるような近似関数を求める



(a) 過去の気温と売上高の関係



(b) 1 次式による回帰

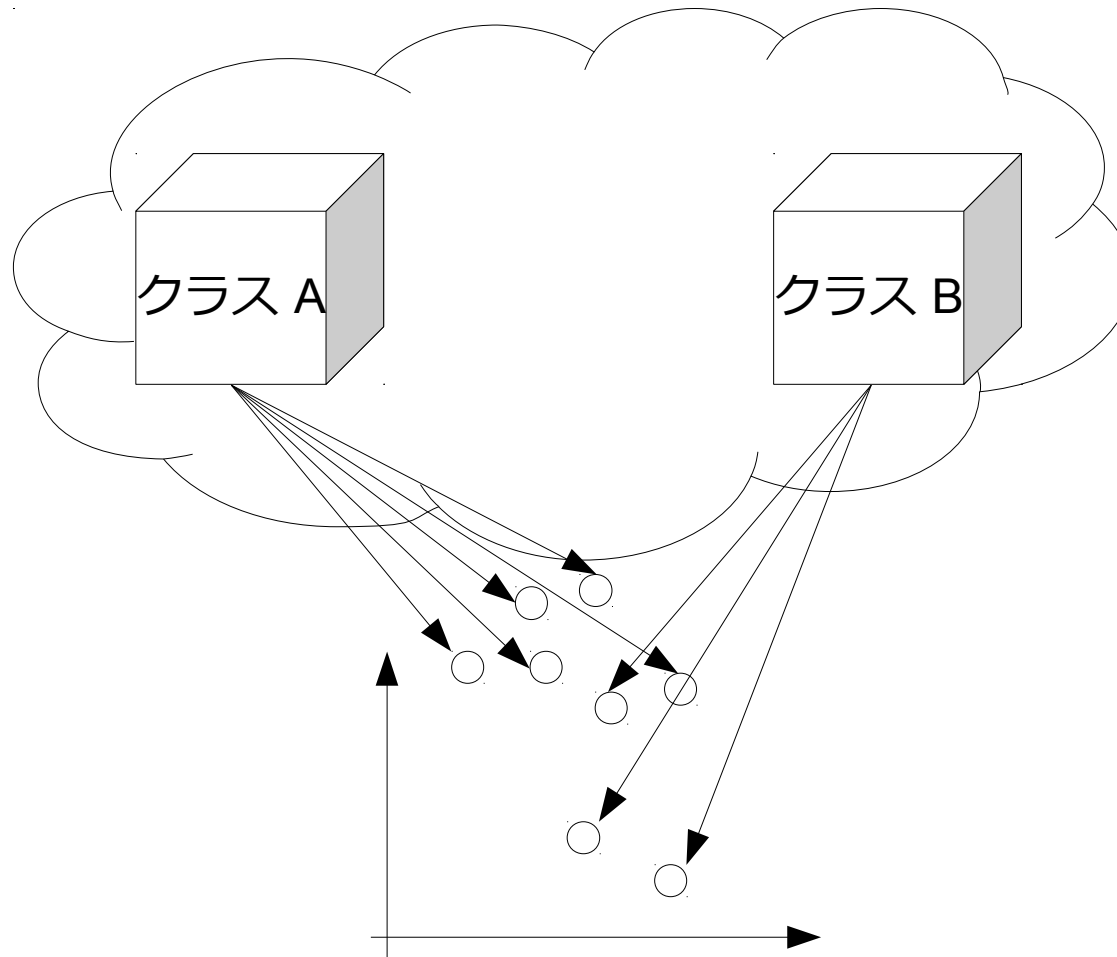


(c) 高次の式による回帰

一般化という視点でどちらが適しているか

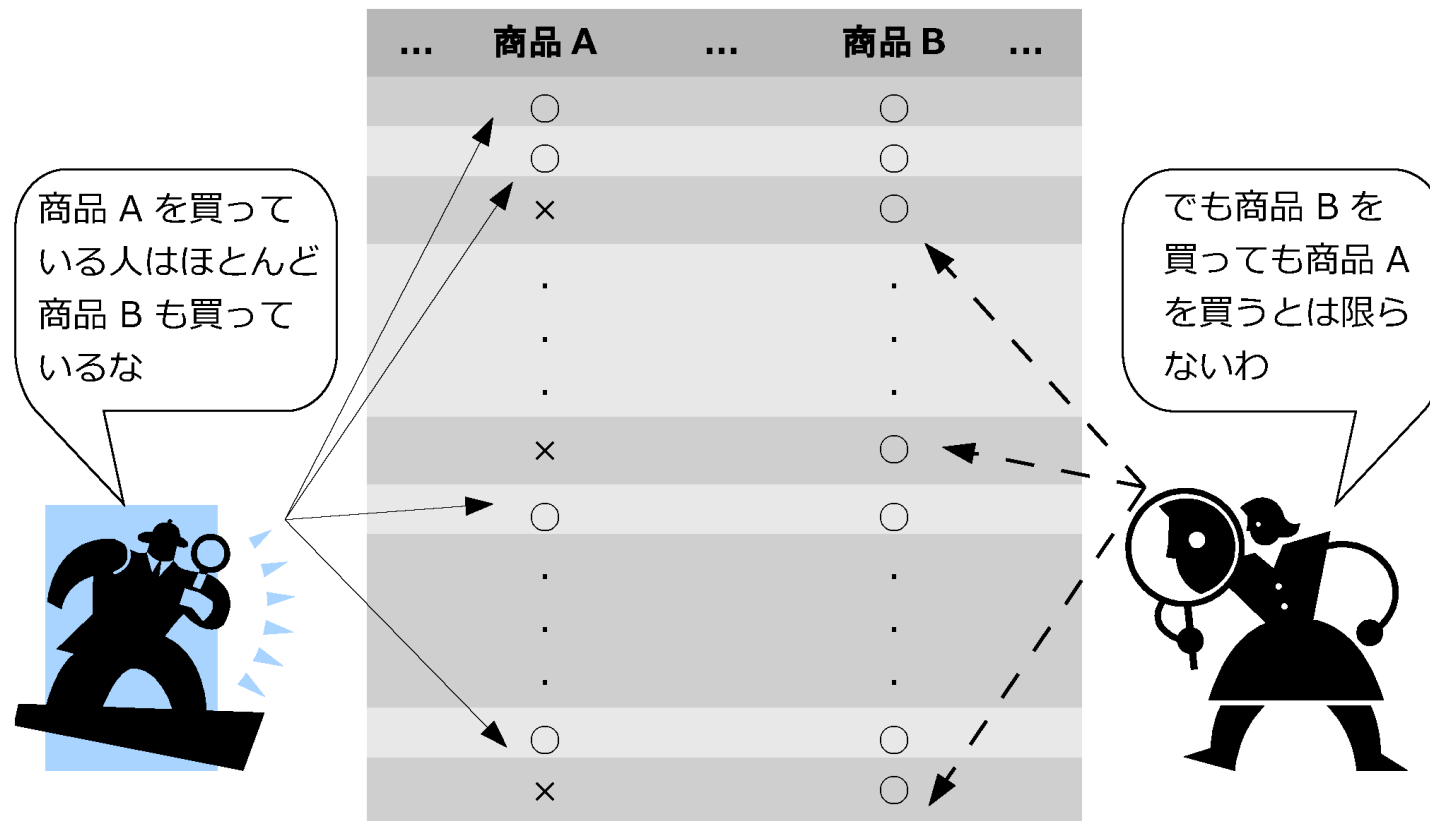
1.3.2 教師なし学習

- モデル推定
 - データを生じさせたクラスを推定



1.3.2 教師なし学習

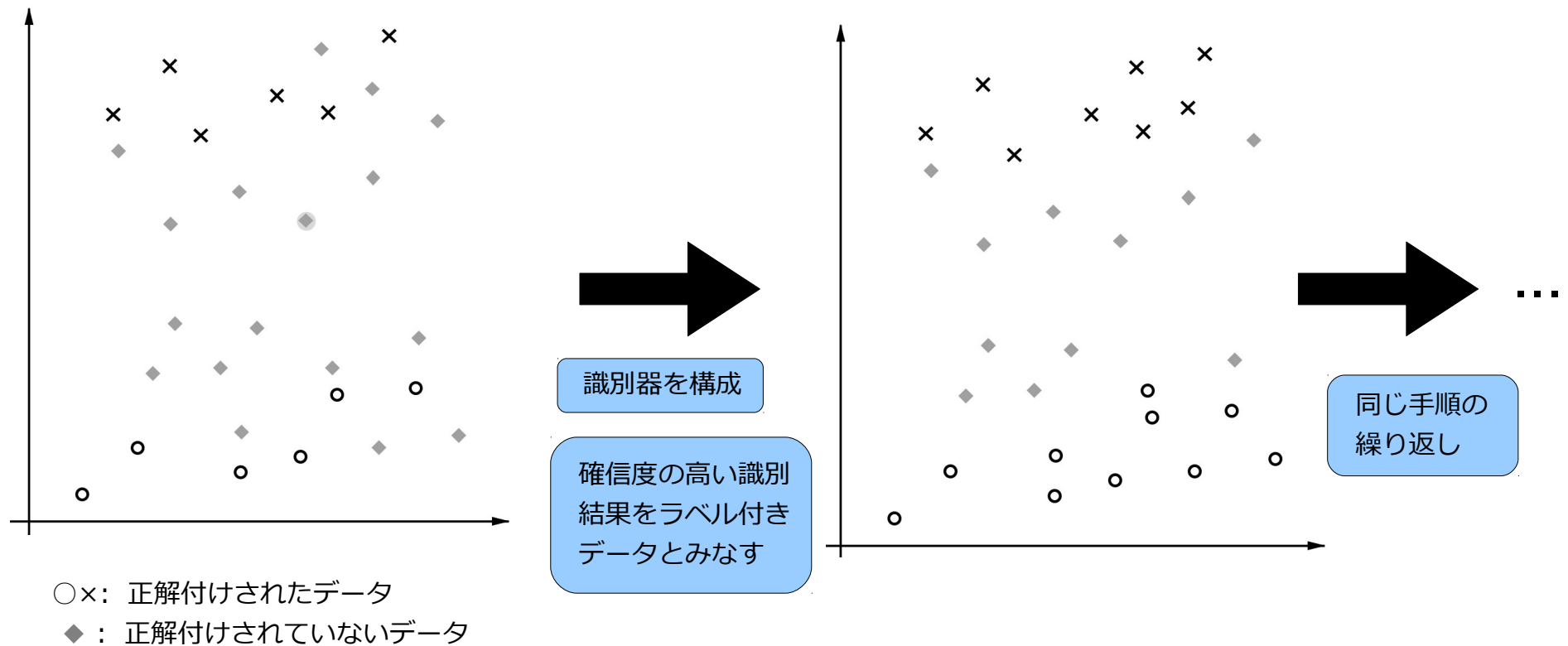
- パターンマイニング
 - 頻出項目や隠れた規則性を発掘



「商品 A を購入」 \Rightarrow (ならば) 「商品 B を購入」

1.3.3 中間的手法

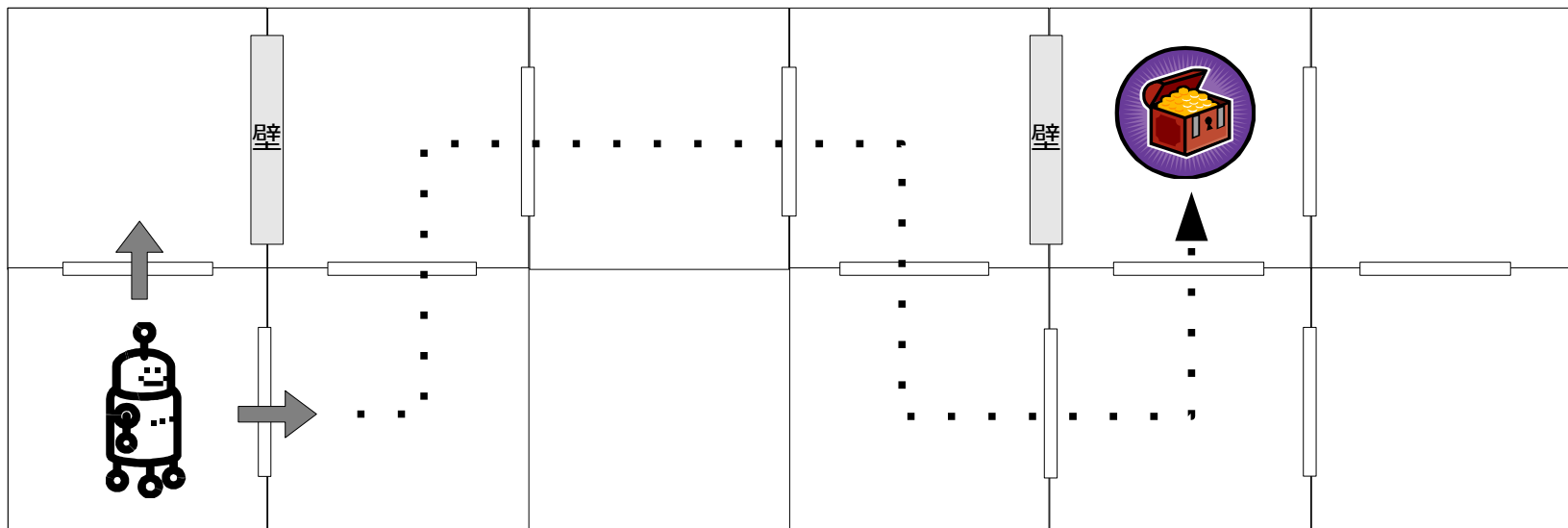
- 半教師あり学習
 - 繰り返しによる学習データの増加



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
 - 教師なし学習で初期値を設定→表現学習
 - 教師あり学習で識別能力を学習

