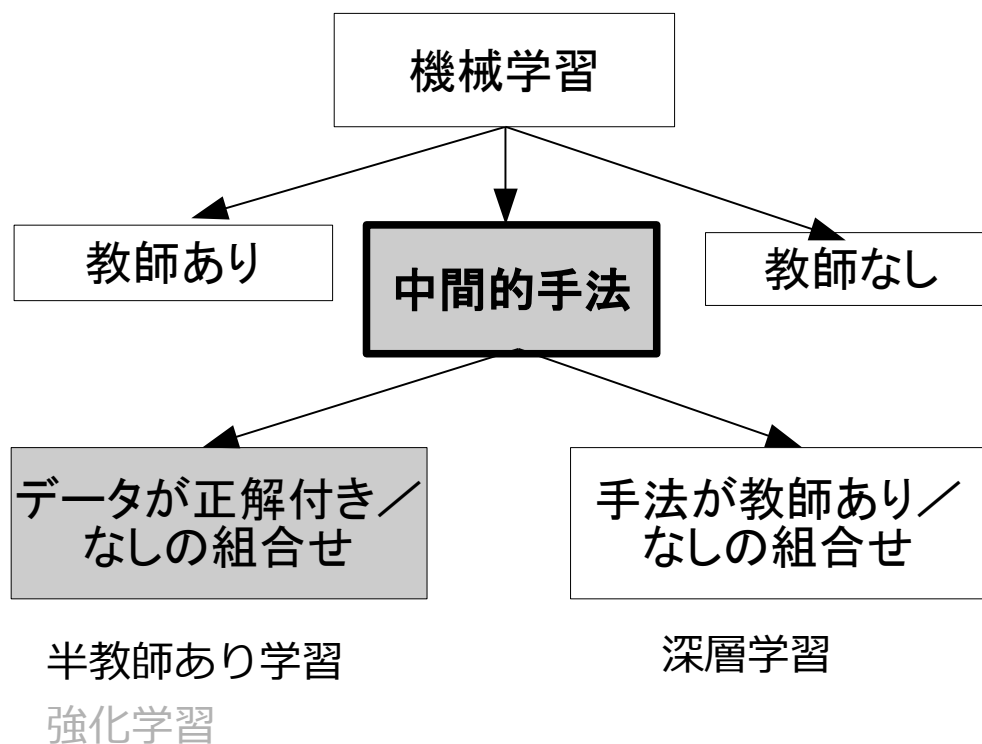


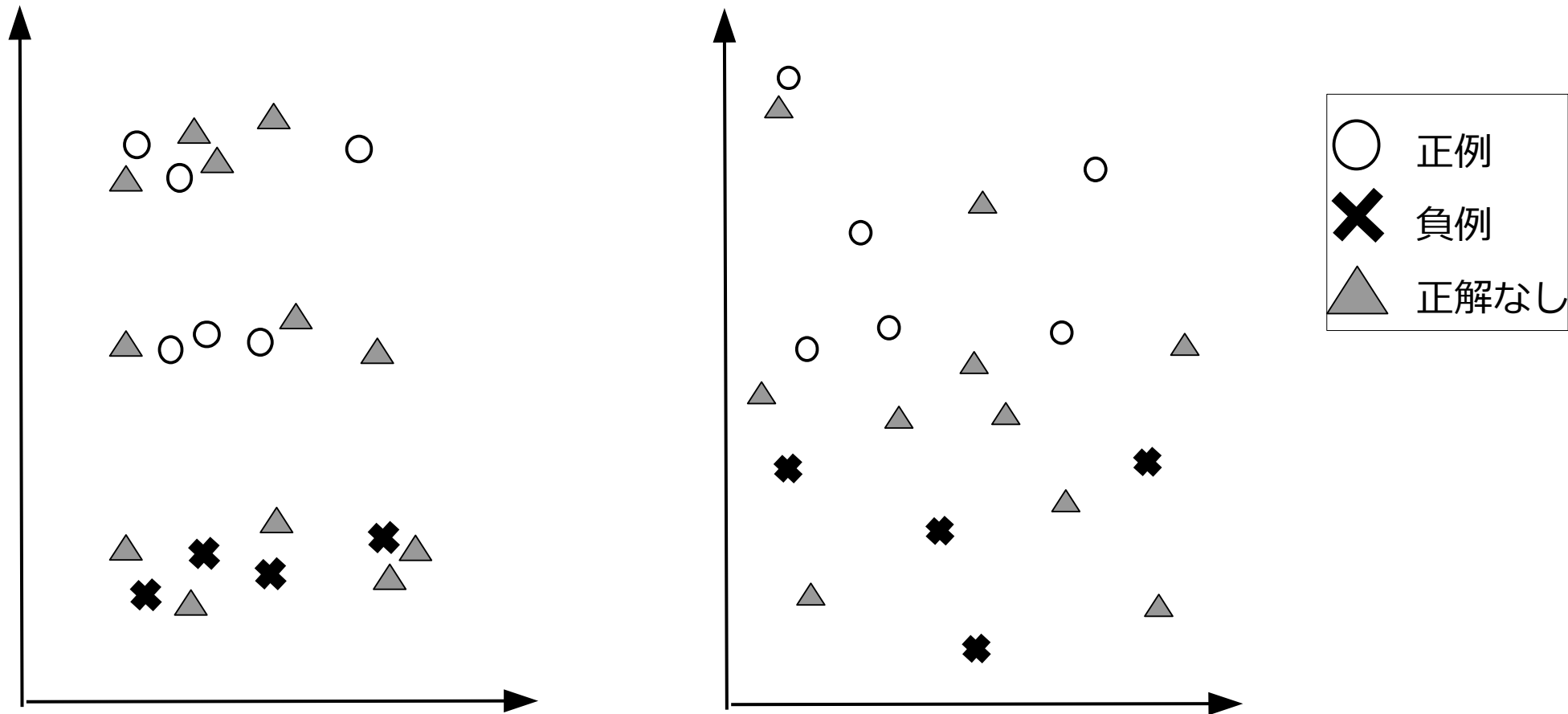
# 13 章 半教師あり学習



# 13.1 半教師あり学習とは

## 13.1.1 数値特徴の場合

- 半教師あり学習に適した数値特徴データの性質



半教師あり学習に適するデータ

半教師あり学習に適さないデータ

## 13.1.1 数値特徴の場合

- 半教師あり学習が可能なデータ
  - 半教師あり平滑性仮定
    - 二つの入力が高密度領域で近ければ、出力も関連している
  - クラスタ仮定
    - もし入力と同じクラスタに属するなら、それらは同じクラスになりやすい
  - 低密度分離
    - 識別境界は低密度領域にある
  - 多様体仮定
    - 高次元のデータは、低次元の多様体上に写像できる

## 13.1.2 ラベル特徴の場合

- オーバーラップ

Positive ○

... よかった。 ..  
...  
高性能 ..  
... 満足

?

...  
...  
高性能 ..  
... 満足 .  
....

?

.....  
...  
高性能 ..  
...  
... よかった。

Negative

×

... 壊れた。 ..  
...  
不満 ..  
...  
... 買わない

?

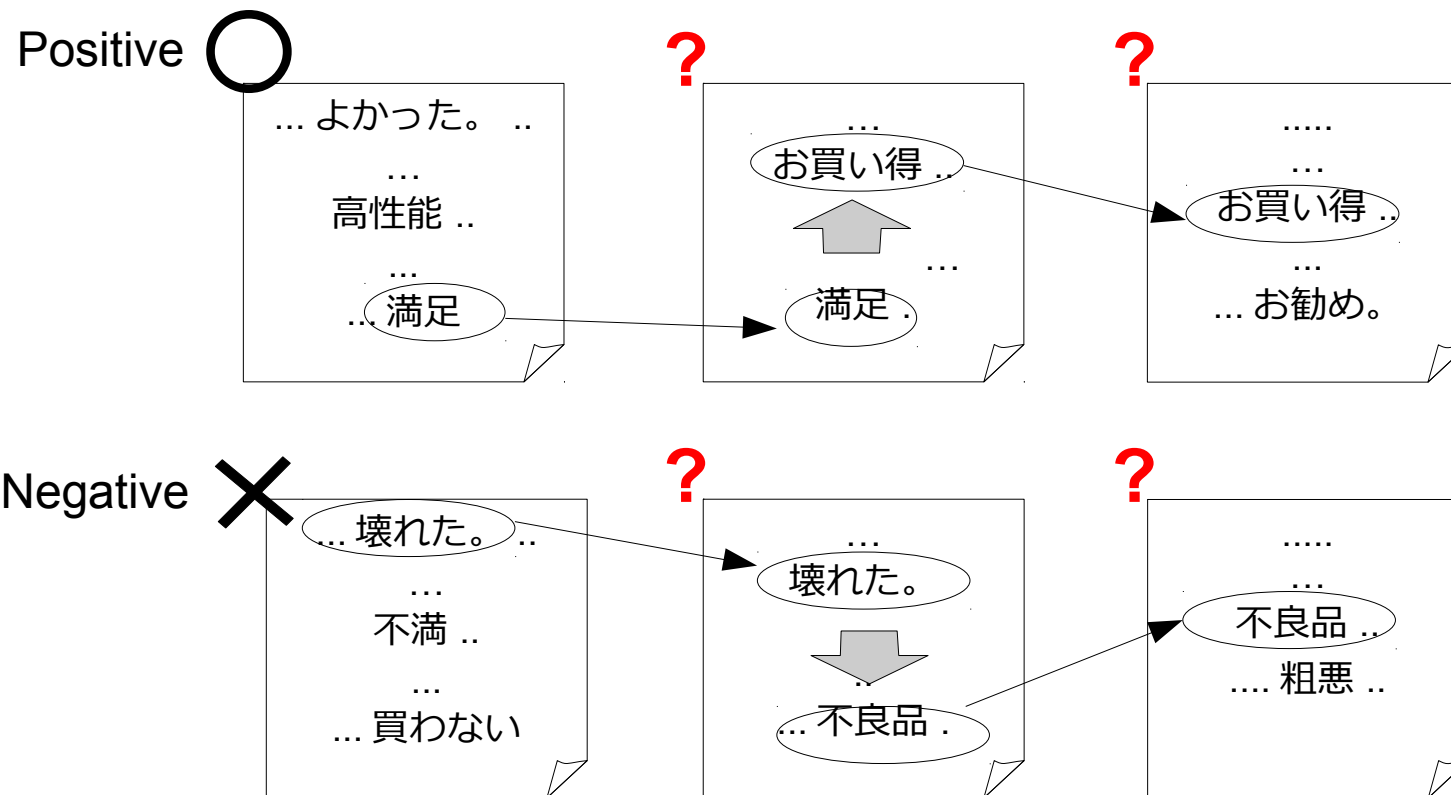
...  
...  
壊れた。 ..  
... 買わない .  
....

?

.....  
...  
不満 ..  
...  
... 買わない

## 13.1.2 ラベル特徴の場合

- 伝播



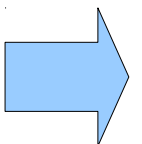
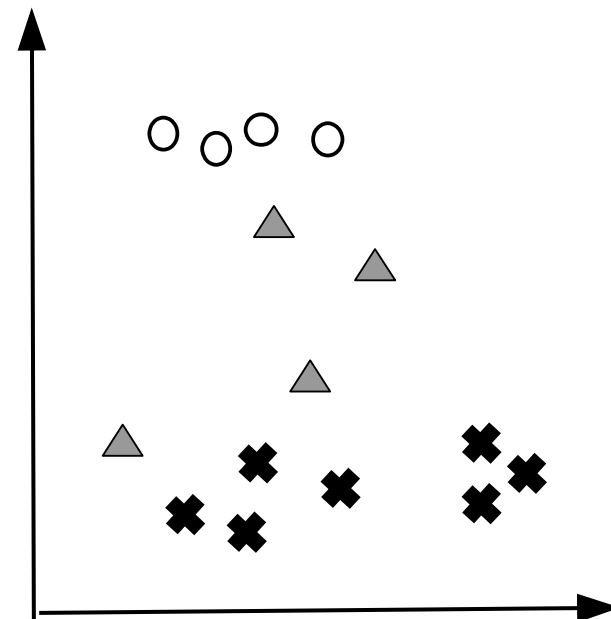
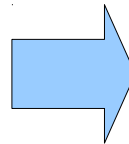
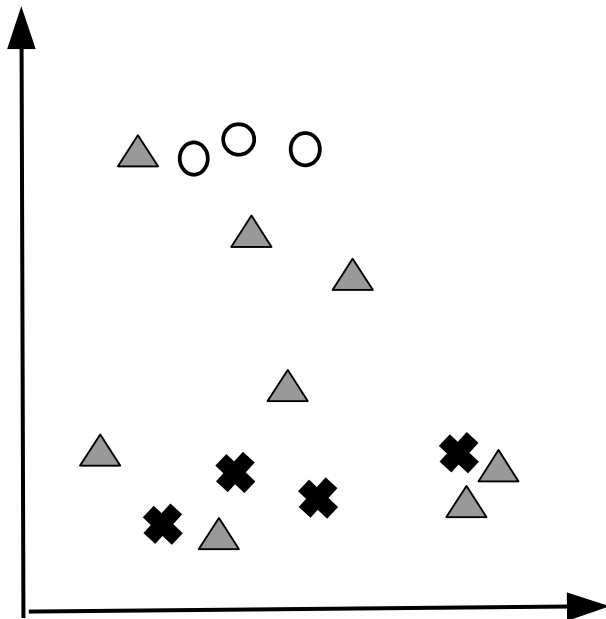
## 13.1.3 半教師あり学習のアルゴリズム

- 半教師あり学習の基本的な考え方
  - 正解付きデータで識別器を作成
  - 正解なしデータで識別器のパラメータを調整
- 識別器に対する要求
  - 確信度の出力：正解なしデータに対する出力を信用するかどうかの判定に必要

## 13.2 自己学習

- 自己学習のアルゴリズム

1. 正解付きデータで初期識別器を作成
2. 正解なしデータの識別結果のうち、確信度の高いものを、正解付きデータとみなす
3. 新しい正解付きデータで、識別器を学習
4. 2, 3 を繰り返す



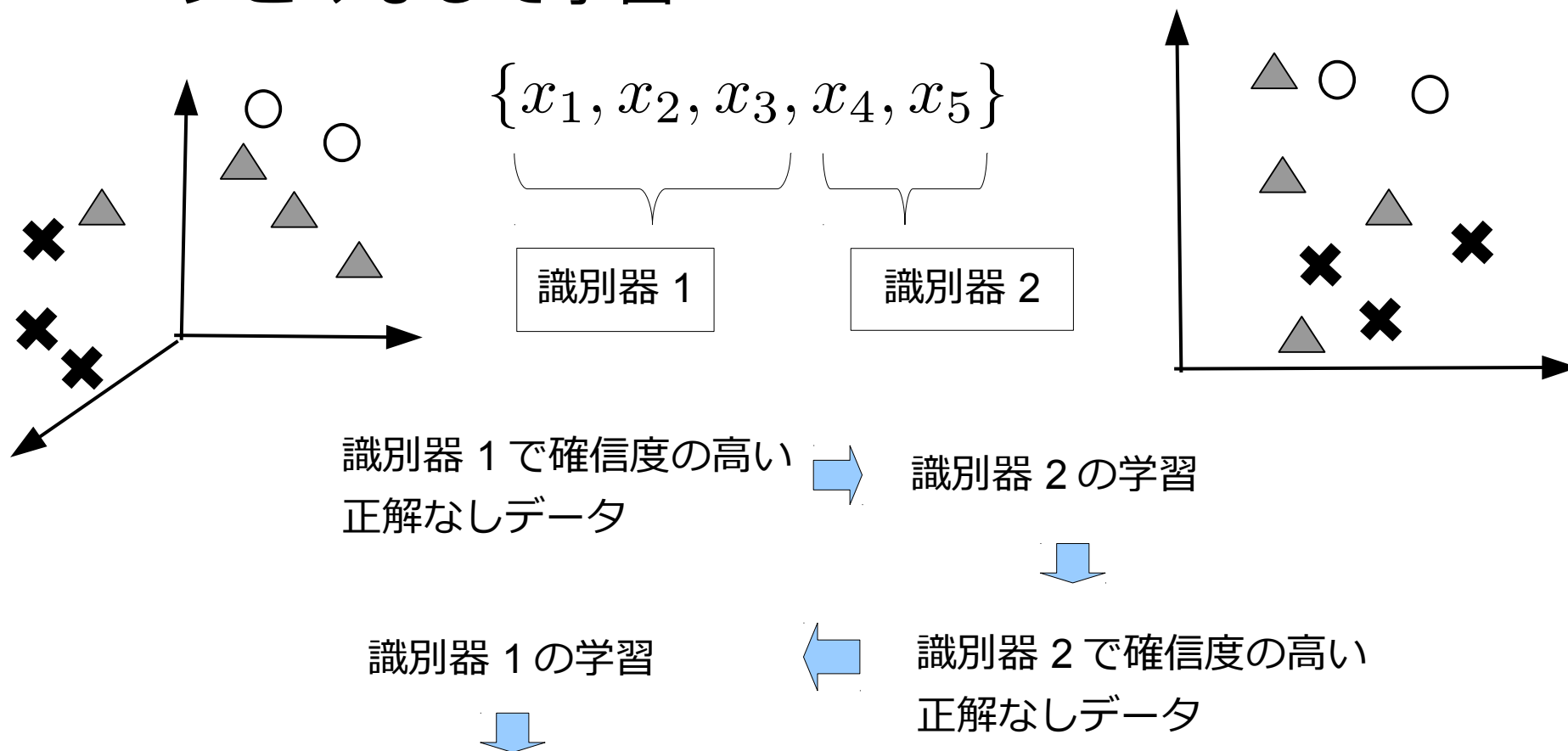
## 13.2 自己学習

- 自己学習の性質
  - クラスタ仮定や低密度分離が満たされるデータに対しては、高い性能が期待できる
  - 低密度分離が満たされていない場合、初期識別器の誤りが拡大してゆく可能性がある



## 13.3 共訓練

- 共訓練とは
  - 判断基準が異なる識別器を交互に用いる
  - 片方の確信度が高いデータを、相手が正解付きデータとみなして学習



## 13.3 共訓練

- 共訓練の特徴
  - 学習初期の誤りに対して頑健
- 共訓練の問題点
  - それぞれが識別空間として機能する特徴集合を、どのようにして作成するか
  - 全ての特徴を用いる識別器よりも高性能な識別器が作成できるか