

機械学習②

クラス分類

東京大学 大学院工学系研究科

吉元俊輔

機械学習に関する用語説明

- 説明（独立）変数・目的（従属）変数

変数 x が、変数 t を用いて数式表現できる時、 t を説明変数、 x を目的変数という

- モデル選択（モデル化，モデル比較）

説明変数と目的変数の関係を数式表現することをモデル選択という

- 入力（観測）変数・出力（目標）変数

モデルにおいて，入力（観測）される変数を入力変数，所望の値を出力変数という

- 学習（訓練）

観測されたデータに基づき，モデルパラメータを最適化することを学習という

- 教師あり・なし

学習において，入出力が与えられるのを教師あり，入力データのみが与えられるのを教師なしという

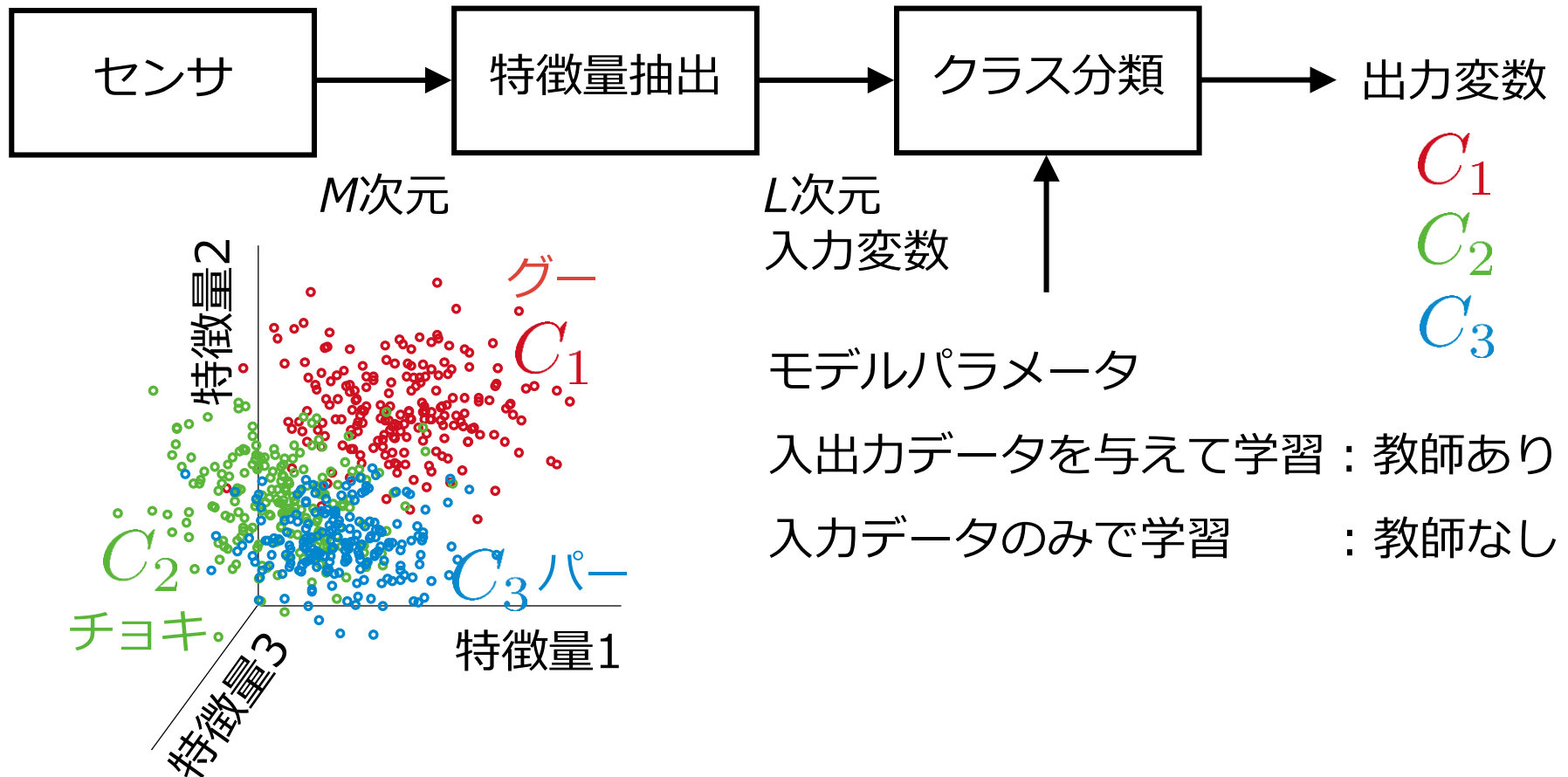
- 回帰・クラス分類

- 入力変数から，連続的な出力変数を求める問題を回帰という

- 入力変数から，有限個のカテゴリに分ける問題をクラス分類という

特徴量を用いたクラス分類の基本

1. 予めモデルパラメータを最適化する = 学習
2. 未知のデータに対して分類を判別する = 予測（認識）



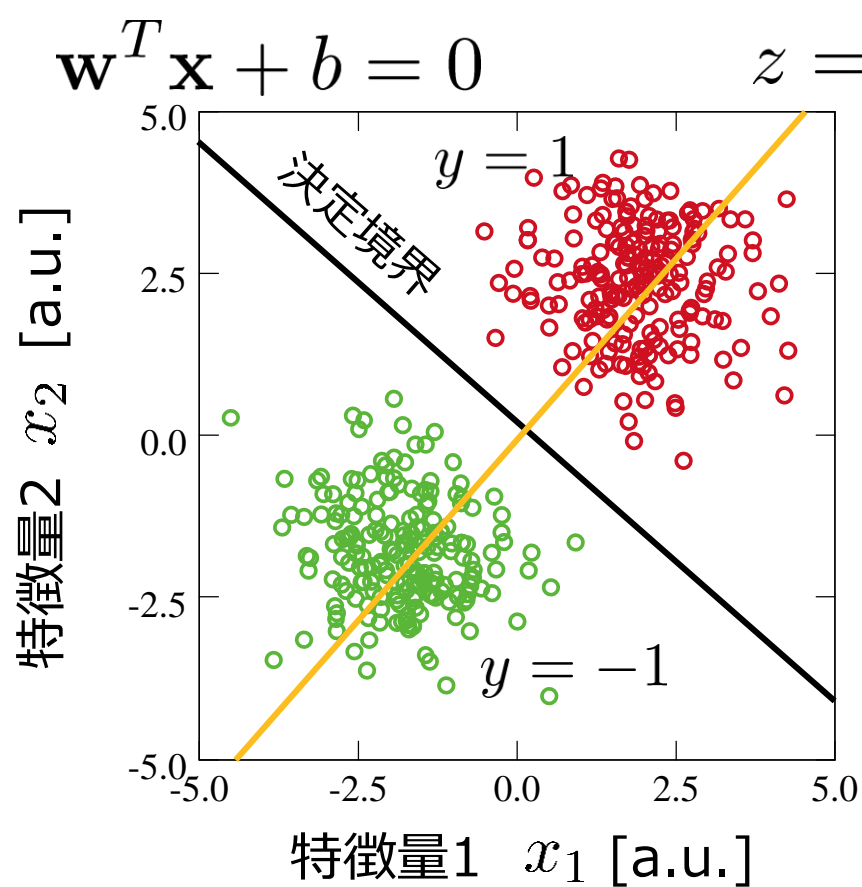
クラス分類の方法

- 判別分析
- 決定木
- ロジスティック回帰
- **Support Vector Machine**
- ランダムフォレスト
- ニューラルネットワーク
- **Deep Neural Network**

線形判別分析

線形判別分析 (Linear Discriminant Analysis)

直線 z に射影した各クラスの分離度が最大になるような
線形変換ベクトル \mathbf{w} を見つけ、決定境界を定める



線形判別器：

$$y = \text{sgn}(\mathbf{w}^T \mathbf{x} + b)$$

特徴ベクトル：

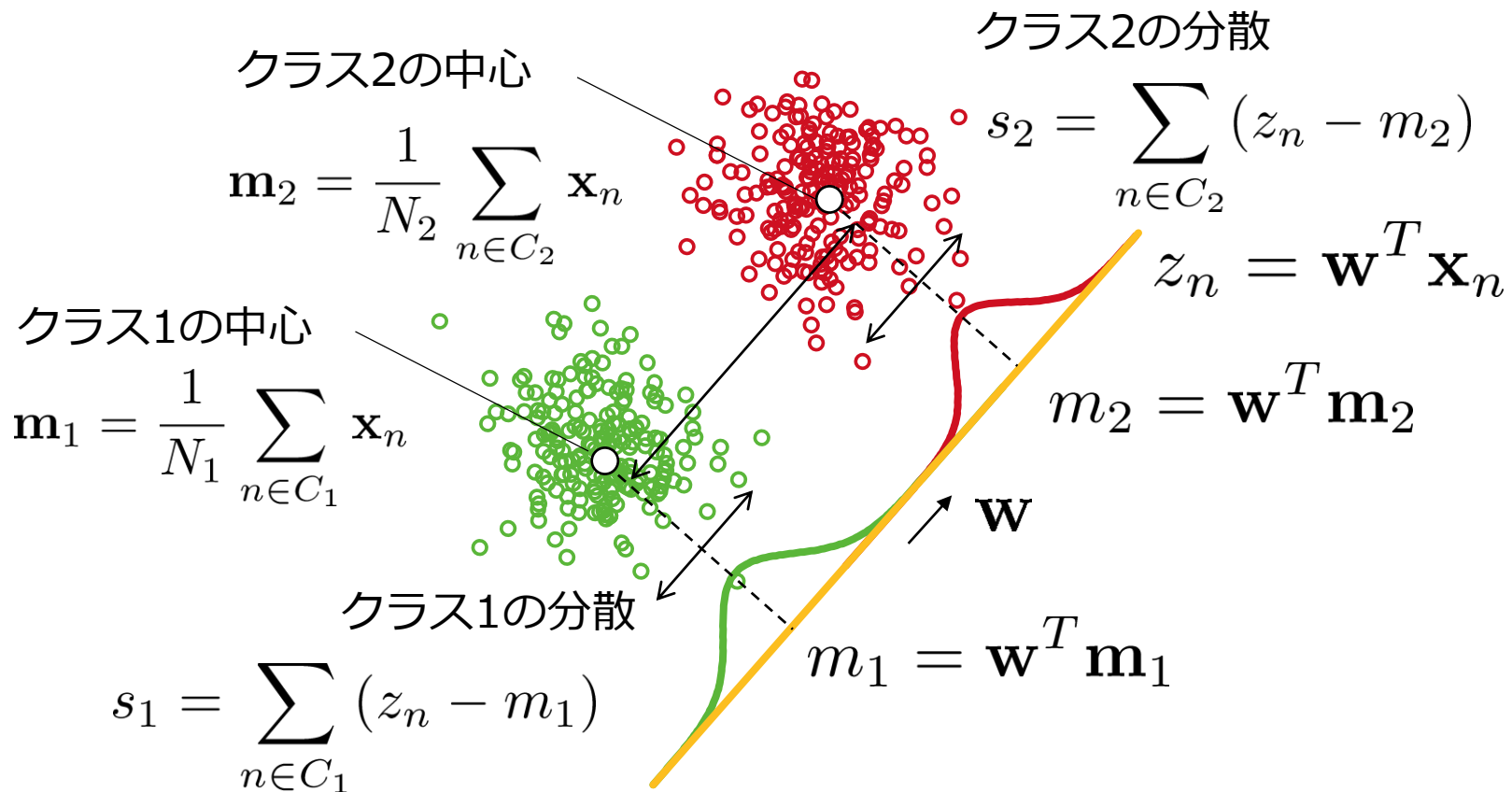
$$\mathbf{x} = (x_1, x_2, \dots, x_L)^T$$

モデルパラメータ：

$$\mathbf{w} = (w_1, w_2, \dots, w_L)^T$$
$$b$$

線形変換ベクトルの決定方針（分離度とは？）

- クラス間の平均の差が最大となる変換を見つける
- 各クラス内の分散が最小となる変換を見つける




最適化問題の定義

クラス間の平均の差が
最大となる変換を見つける


$$\begin{aligned} S'_B &= (m_2 - m_1)^2 \\ &= ((\mathbf{m}_2 - \mathbf{m}_1)^T \mathbf{w})^2 \\ &= \mathbf{w}^T (\mathbf{m}_2 - \mathbf{m}_1) (\mathbf{m}_2 - \mathbf{m}_1)^T \mathbf{w} \\ &= \mathbf{w}^T \mathbf{S}_B \mathbf{w} \end{aligned}$$

各クラス内の分散が
最小となる変換を見つける

$$\begin{aligned} S'_w &= \sum_{n \in C_1} (z_n - m_1)^2 + \sum_{n \in C_2} (z_n - m_2)^2 \\ &= \mathbf{w}^T \sum_{k=1,2} \sum_{n \in C_k} (\mathbf{x}_n - \mathbf{m}_k) (\mathbf{x}_n - \mathbf{m}_k)^T \mathbf{w} \\ &= \mathbf{w}^T \mathbf{S}_w \mathbf{w} \end{aligned}$$


$$J(\mathbf{w}) = \frac{S'_B}{S'_w} = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}$$

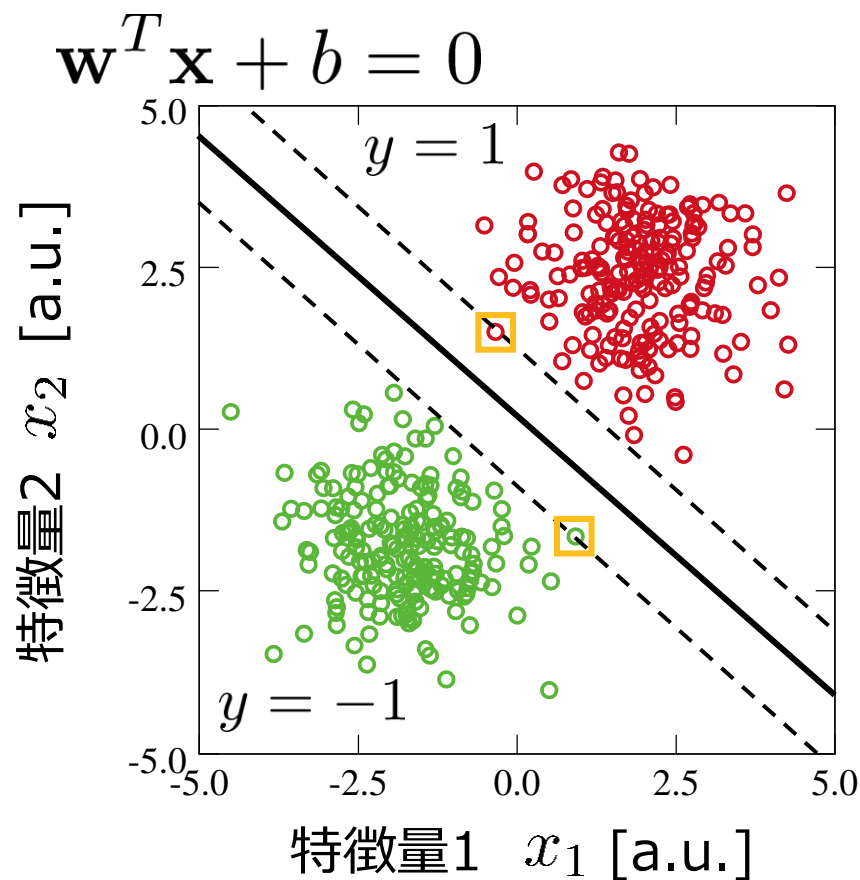
が最大となる \mathbf{w} をを見つける


$$\mathbf{w}_{\text{opt}} \propto \mathbf{S}_W^{-1} (\mathbf{m}_2 - \mathbf{m}_1)$$

Support Vector Machine

Support Vector Machine

データを判別に重要なデータ（サポートベクトル）とそうでないデータに分け，サポートベクトルで判別規則を構築する



線形判別器：

$$y = \text{sgn}(\mathbf{w}^T \mathbf{x} + b)$$

特徴ベクトル：

$$\mathbf{x} = (x_1, x_2, \dots, x_L)^T$$

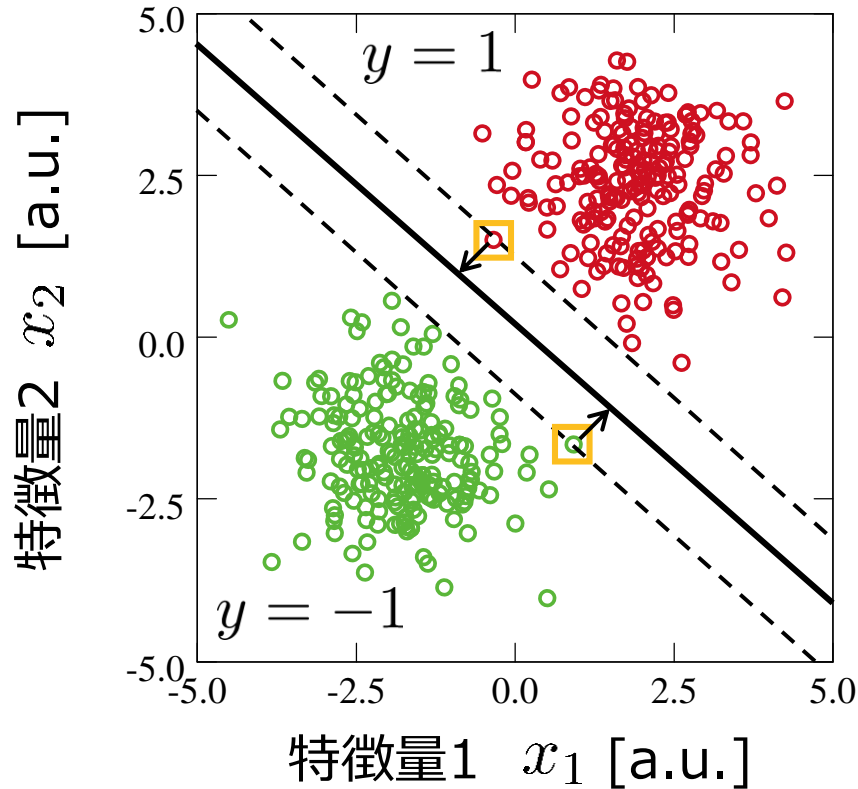
モデルパラメータ：

$$\mathbf{w} = (w_1, w_2, \dots, w_L)^T$$
$$b$$

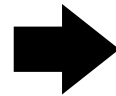
(ハード) マージン最大化

- マージン：各クラスの特徴ベクトルの距離
- 特徴ベクトル \mathbf{x}_n と超平面 $\mathbf{w}^T \mathbf{x} + b = 0$ の距離を最大化

$$\mathbf{w}^T \mathbf{x} + b = 0$$



$$\max_{\mathbf{w}, b} \min_n \frac{|\mathbf{w}^T \mathbf{x}_n + b|}{\|\mathbf{w}\|}$$

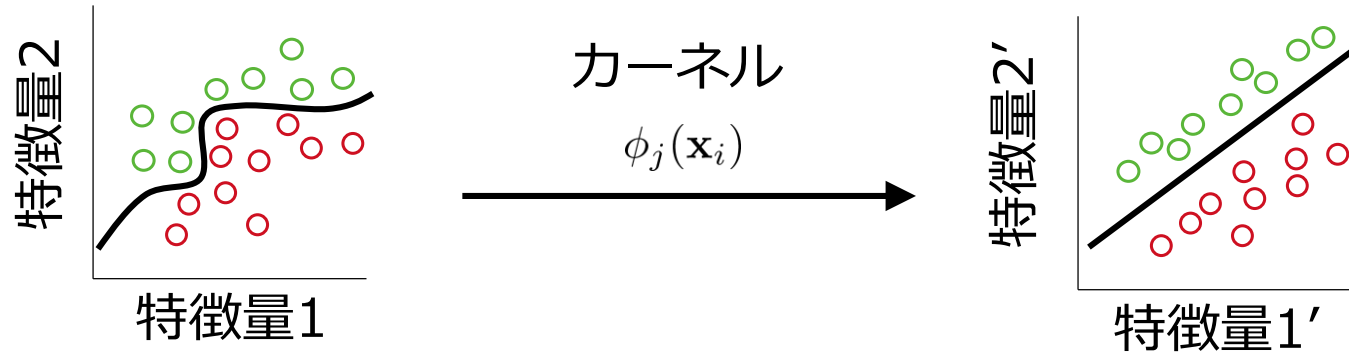


$$\min_{\mathbf{w}, b} \|\mathbf{w}\|^2$$

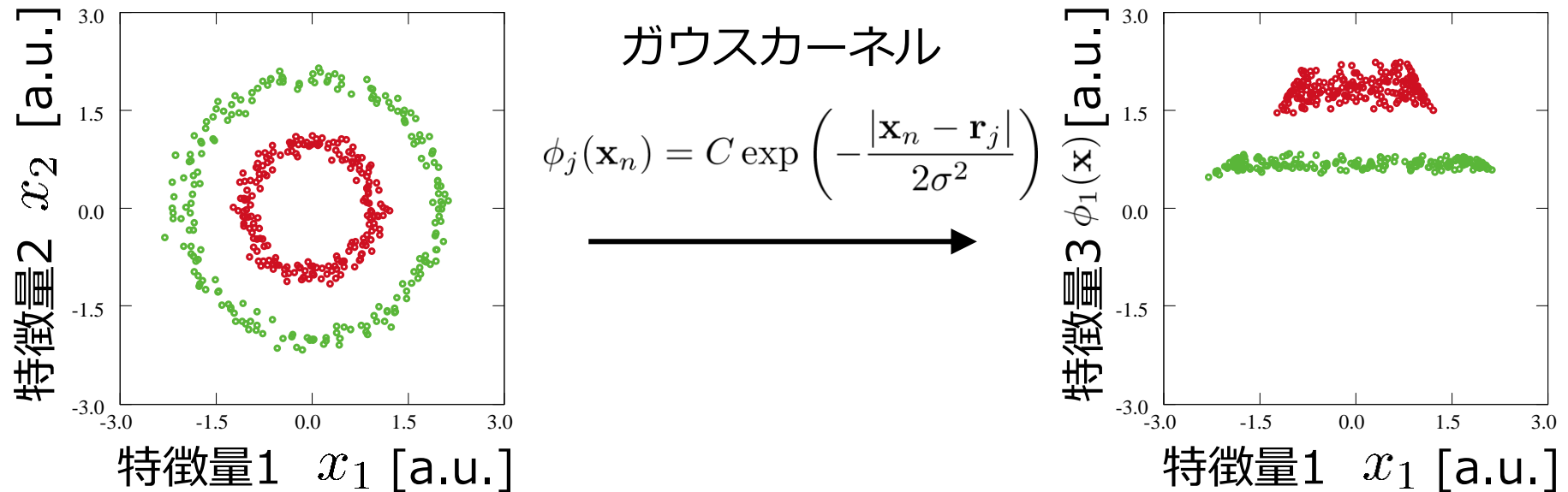
$$s.t. \ y_n (\mathbf{w}^T \mathbf{x}_n + b) \geq 1$$

超平面で判別できない場合

- 判別できる空間に変数変換により写像する



- 特徴ベクトルの次元を増やす



学習と評価

教師あり学習

- 特徴ベクトルとクラスの組み合わせが既知なデータを用意する
- データについて
 - データ数は求めるパラメータの数よりも十分多いことが望ましい
 - データは概ね正規分布に従う事が前提
 - 特徴ベクトル空間をまんべんなくサンプリングしたデータを与える
- 判別性能について
 - 学習に用いたデータで正しく判別できる割合 (=認識率)が用いられる
(求めた識別器を用いて特徴ベクトルを判別し, 正解と比較)
 - 超平面で判別できないような分布のデータには特徴量変換が必要

評価

- 学習に用いたデータに対して，未知の特徴ベクトルが与えられた際に，どの程度正しく識別できるか？
- 交差検証
 - データセットを学習用と評価用に分けて識別率を求める
 - n-fold cross validation法

データセットをn分割し，n回のモデル構築と評価を行い，平均を算出
 - Leave one out法

データセットのうち一つを評価に用いて残りを学習に利用，評価に用いるデータを順番に変えて認識率をそれぞれ求め，平均を算出