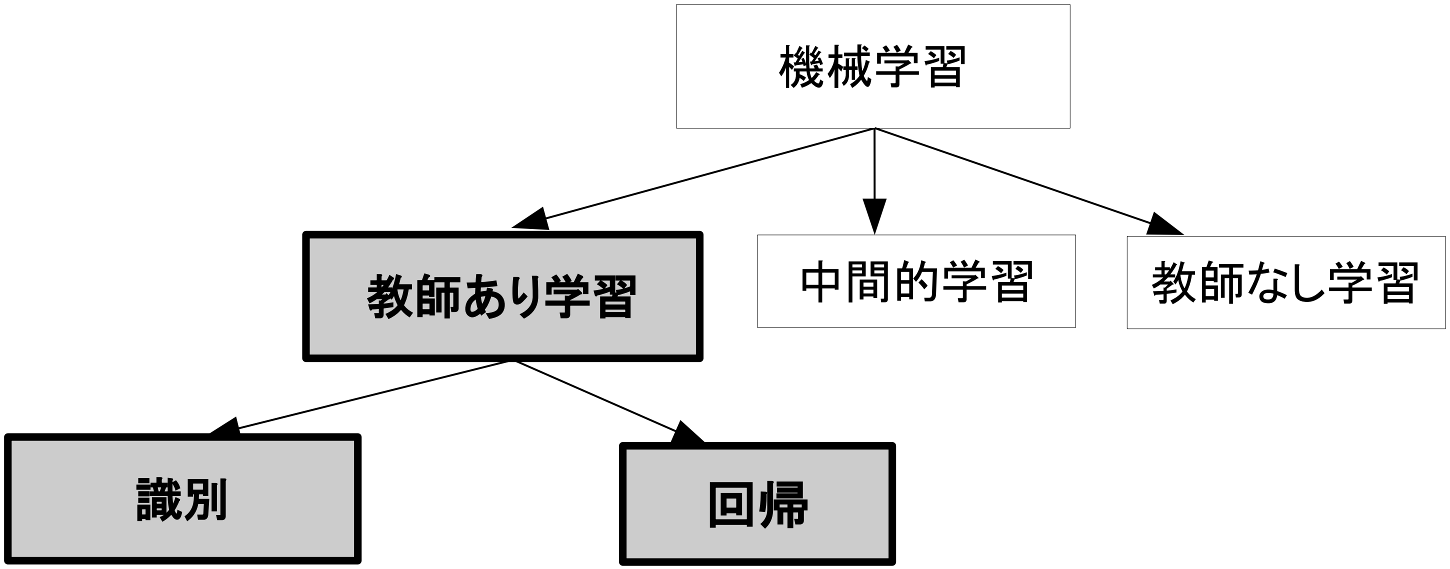
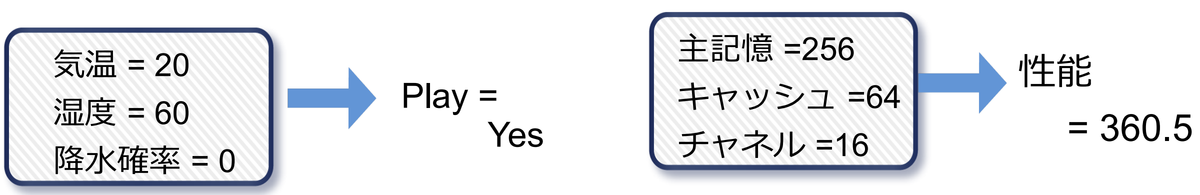
7. サポートベクトルマシン



- 数値特徴

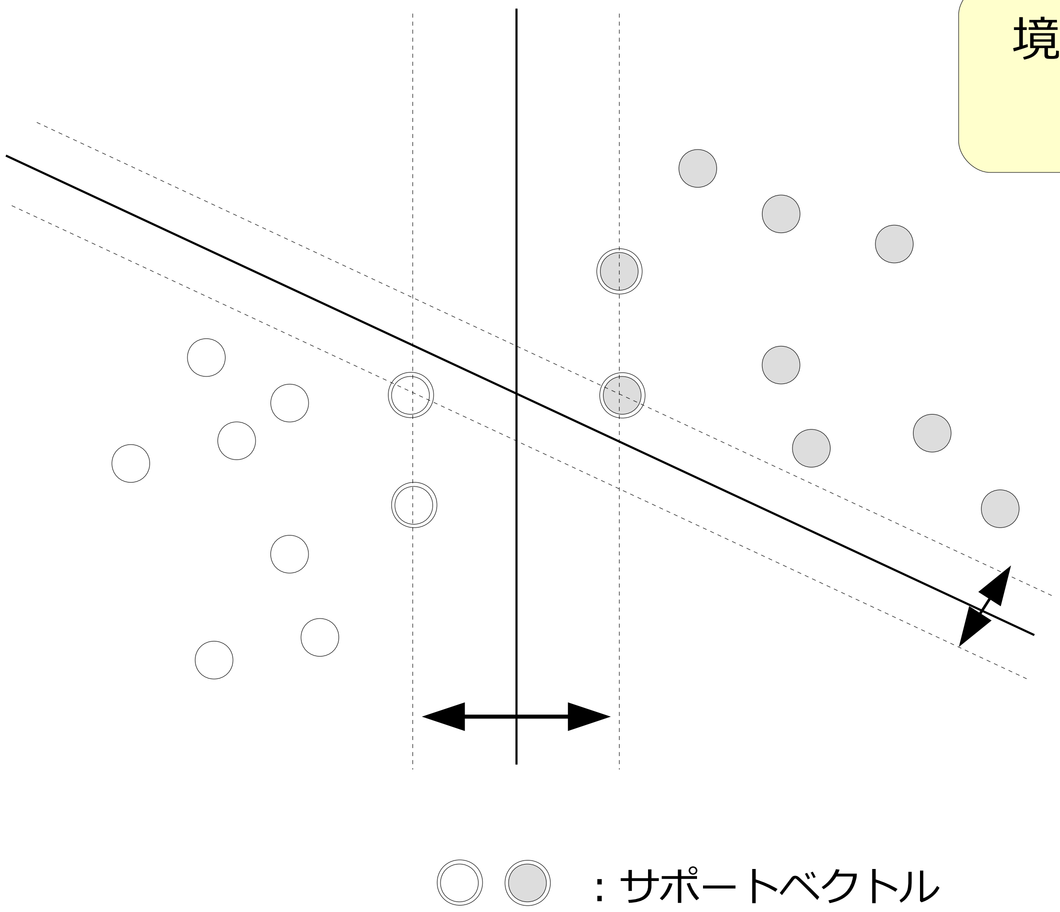


7. サポートベクトルマシン

- マージンを最大化する識別面を求める

識別面と、最も近いデータとの距離

境界部分のデータにのみ注目



7.1 サポートベクトルマシンとは

- 学習データ

$$\chi = \{(\boldsymbol{x}_i, y_i)\} \quad i = 1, \dots, N, \quad y_i = 1 \ or \ -1$$

- 識別面の式

$$\boldsymbol{w}^T \boldsymbol{x}_i + w_0 = 0$$

- 識別面の制約（係数を定数倍しても平面は不変）

$$\min_{i=1,\dots,N} |\boldsymbol{w}^T \boldsymbol{x}_i + w_0| = 1$$

点と直線の距離の公式

$$r = \frac{|ax + by + c|}{\sqrt{a^2 + b^2}}$$

- 学習パターンと超平面との最小距離

$$\min_{i=1,\dots,N} Dist(\boldsymbol{x}_i) = \min_{i=1,\dots,N} \frac{|\boldsymbol{w}^T \boldsymbol{x}_i + w_0|}{||\boldsymbol{w}||} = \frac{1}{||\boldsymbol{w}||}$$

7.1 サポートベクトルマシンとは

- 目的関数： $\min \frac{1}{2}||\boldsymbol{w}||^2$
- 制約条件： $y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) \geq 1 \quad i = 1, \ldots, N$
- 解法：ラグランジュの未定乗数法
  - 問題 $\min f(x) \quad s.t.\ g(x) = 0$
  - ラグランジュ関数 $L(x, \alpha) = f(x) + \alpha g(x)$
    - $\alpha \geqq 0$
    - x, α で偏微分して 0 になる値が極値

7.1 サポートベクトルマシンとは

- 計算

$$L(\boldsymbol{w}, w_0 . \alpha) = \frac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{N} \alpha_i (y_i (\boldsymbol{w}^T \boldsymbol{x} + w_0) - 1)$$

$$\frac{\partial L}{\partial w_0} = 0 \quad \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \boldsymbol{w}} = 0 \quad \Rightarrow \boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x}_i$$

$$L(\alpha) = \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j - \sum_{i=1}^{N} \alpha_i$$

α についての
2 次計画問題

7.1 サポートベクトルマシンとは

- 定数項の計算
  - 各クラスのサポートベクトルから求める

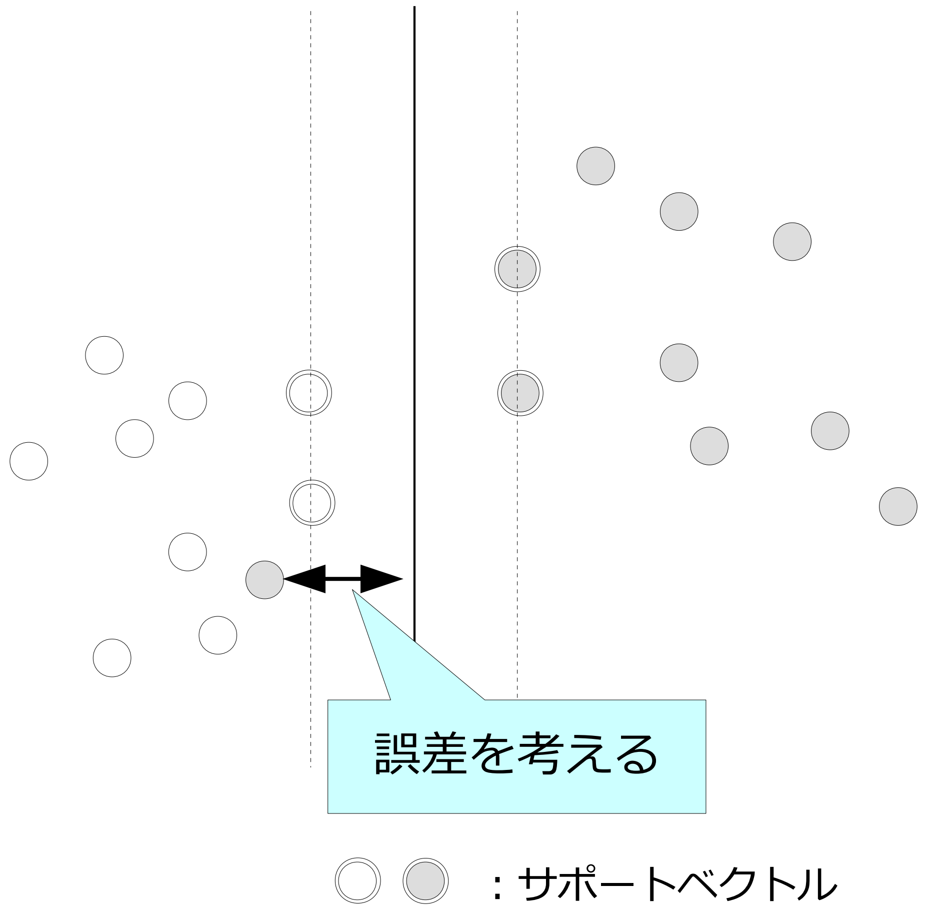
$$w_0 = -\frac{1}{2}(\boldsymbol{w}^T\boldsymbol{x}_{s1} + \boldsymbol{w}^T\boldsymbol{x}_{s2})$$

- 識別関数

$$\begin{aligned} g(\boldsymbol{x}) &= \boldsymbol{w}^T\boldsymbol{x} + w_0 \\ &= \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}^T\boldsymbol{x}_i + w_0 \end{aligned}$$

7.2 ソフトマージンによる誤識別データの吸収

- 少量のデータが線形分離性を妨げている場合



7.2 ソフトマージンによる誤識別データの吸収

- スラック変数 ξ_i の導入

$$y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) \geq 1 - \xi_i \quad i = 1, \ldots, N$$

- 最小化問題の修正

$$\min(\frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{N}\xi_i)$$

スラック変数も
小さい方がよい

- 計算結果
  - α_i の 2 次計画問題に $0 \leq \alpha_i \leq C$ が加わるだけ

7.2 ソフトマージンによる誤識別データの吸収

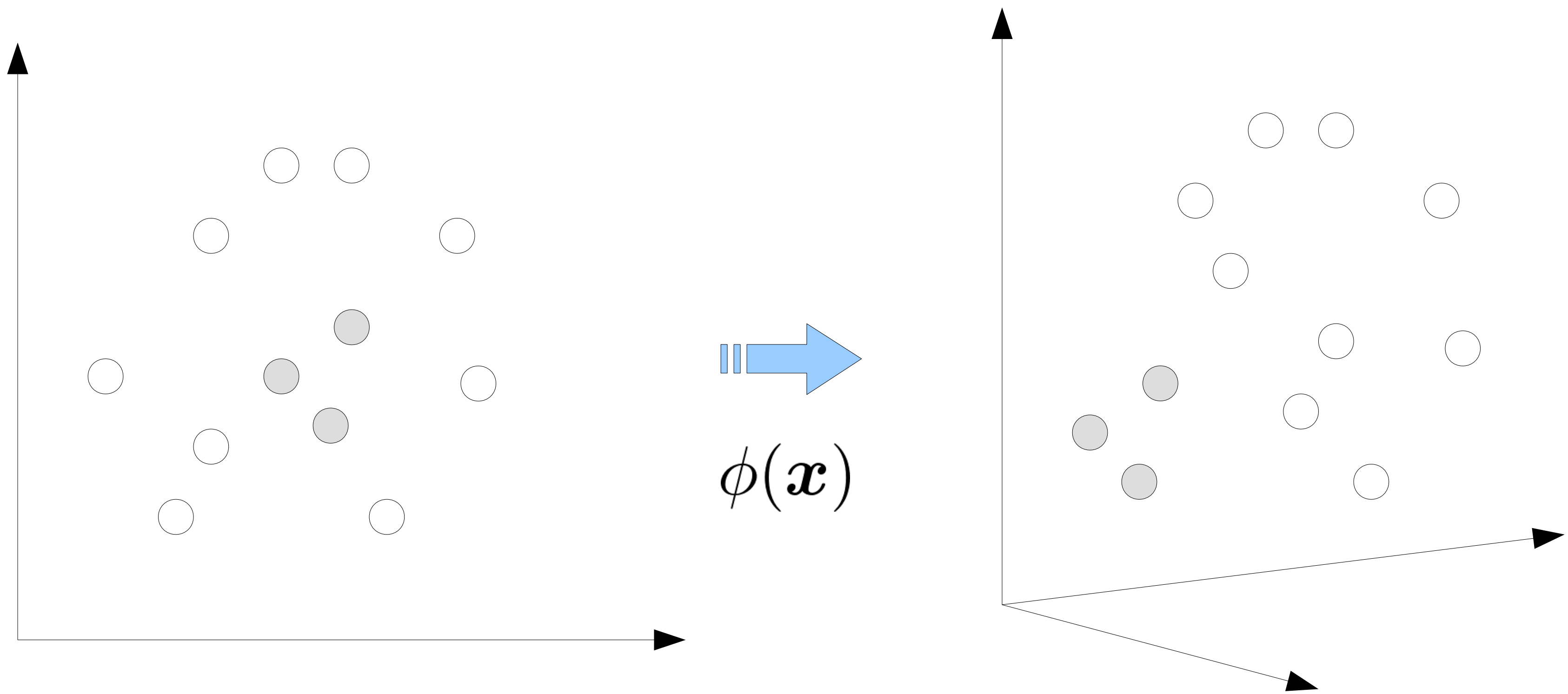
- C: エラー事例に対するペナルティ
  - 大きな値：誤識別データの影響が大きい

    → 複雑な識別面
  - 小さな値：誤識別データの影響が小さい

    → 単純な識別面

7.3 カーネル関数を用いた SVM

- 特徴ベクトルの次元を増やす



2 次元で線形分離不可能なデータ

線形分離可能な 3 次元へ

ただし、元の空間でのデータ間の
距離関係は保持するように

7.3 カーネル関数を用いた SVM

- 非線形変換関数： $\phi(\boldsymbol{x})$
- カーネル関数

$$K(\boldsymbol{x}, \boldsymbol{x}') = \phi(\boldsymbol{x})^T \phi(\boldsymbol{x}')$$

2 つの引数値の
近さを表す

  - 元の空間での距離が変換後の空間の内積に対応
  - $\boldsymbol{x}$ と $\boldsymbol{x}'$ が近ければ $K(\boldsymbol{x}, \boldsymbol{x}')$ は大きい値

7.3 カーネル関数を用いた SVM

- カーネル関数の例（ scikit-learn の定義）
  - 線形　$K(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^T \boldsymbol{x}'$
    - 元の特徴空間でマージン最大の平面
  - 多項式　$K(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}^T \boldsymbol{x}' + r)^d$
    - d 項の相関を加える
  - RBF　$K(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\gamma ||\boldsymbol{x} - \boldsymbol{x}'||^2)$
    - γ の値：大→複雑　小→単純な識別面
  - シグモイド　$K(\boldsymbol{x}, \boldsymbol{x}') = \tanh(\boldsymbol{x}^T \boldsymbol{x}' + r)$
    - ベクトルの近さを基準に閾値関数的な振る舞い

7.3 カーネル関数を用いた SVM

- 変換後の識別関数：　$g(\boldsymbol{x}) = \boldsymbol{w}^T \phi(\boldsymbol{x}) + w_0$
- SVM で求めた $\boldsymbol{w}$ の値を代入

$$
\begin{aligned}
g(\boldsymbol{x}) &= \sum_{i=1}^{N} \alpha_i y_i \phi(\boldsymbol{x})^T \phi(\boldsymbol{x}_i) + w_0 \\
&= \sum_{i=1}^{N} \alpha_i y_i K(\boldsymbol{x}, \boldsymbol{x}_i) + w_0
\end{aligned}
$$

非線形変換の
式は不要！！！

カーネルトリック

まとめ

- Weka デモ
  - Reuters-Corn データ
  - SMO (Poly カーネル、 RBF カーネル)
- SVM
  - マージン最大となる線形識別面を求める方法
  - カーネル関数を用いてデータを高次元空間に写像して線形分離可能性を高める
  - 高次元特徴でも学習が可能