20. サポートベクターマシン

- サポートベクターマシンはデータを2つのクラスに分類する.
- データ
  - i番目のデータベクトル $\boldsymbol{x}(i)$
  - $\boldsymbol{x}(i)$ が属するクラスは $y(i)$によって表される.

$$y(i) = \begin{cases} 1, & \boldsymbol{x}(i) \text{ belongs to class 1,} \\ -1, & \boldsymbol{x}(i) \text{ belongs to class 0.} \end{cases}$$

前と同様に, データを分類するために $\boldsymbol{w}$ を用いる.

$\boldsymbol{w}^T\boldsymbol{x}(i) - b > 0$ ならば $\boldsymbol{x}(i)$ はクラス1に属する.

$\boldsymbol{w}^T\boldsymbol{x}(i) - b < 0$ ならば $\boldsymbol{x}(i)$ はクラス0に属する.

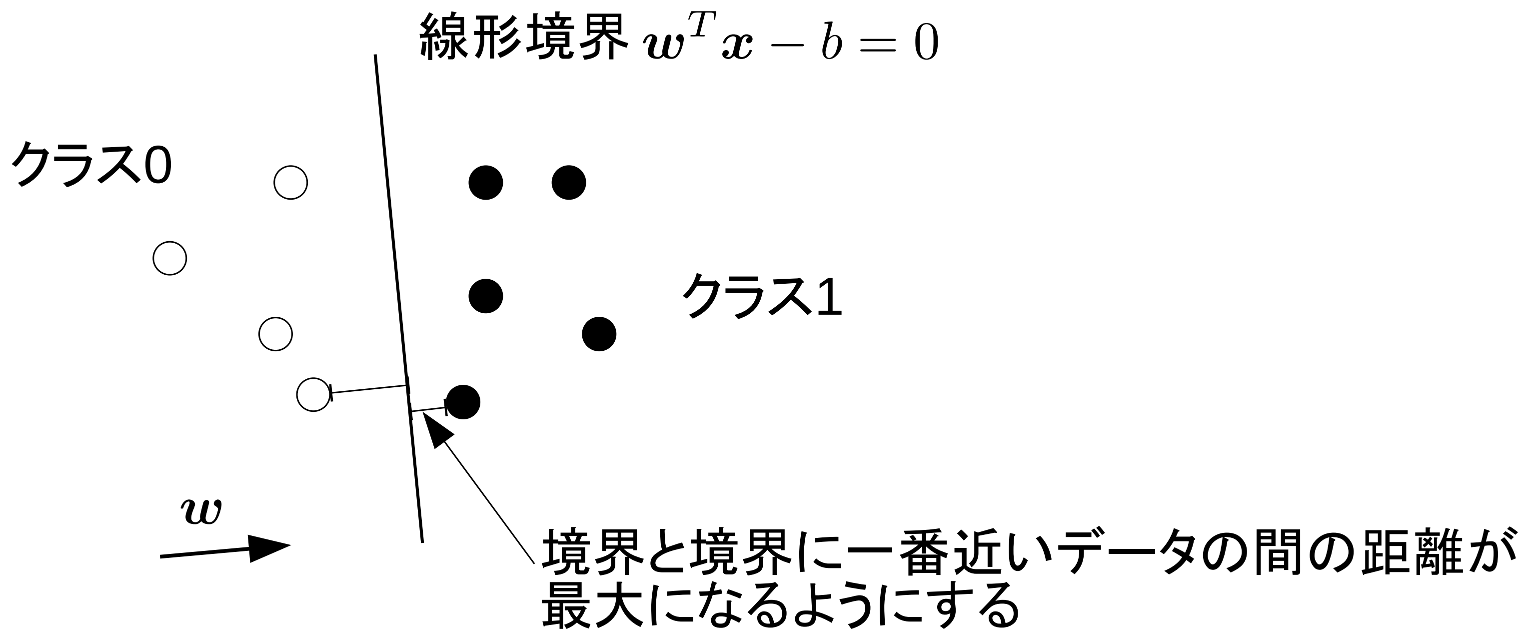
$$\Updownarrow$$

$$y(i)(\boldsymbol{w}^T\boldsymbol{x}(i) - b) > 0$$

良い $\boldsymbol{w}$ と $\boldsymbol{b}$ を求めよう.

20. サポートベクターマシン

- 良い $\boldsymbol{w}$ と b とは?
  - データが境界から十分離れていてほしい



境界とデータ $\boldsymbol{x}(i)$ の間の距離は

$$\frac{|\boldsymbol{w}^T\boldsymbol{x}(i) - b|}{\|\boldsymbol{w}\|},$$

であるので, $\min_i \frac{|\boldsymbol{w}^T\boldsymbol{x}(i) - b|}{\|\boldsymbol{w}\|}$ を最大にしたい.

20. サポートベクターマシン

境界は次式で表される

$$\boldsymbol{w}^T \boldsymbol{x} - b = 0,$$

これは，ゼロではない任意の k を用いて

$$(k\boldsymbol{w})^T \boldsymbol{x} - (kb) = 0,$$

と表しても同じである．

そこで， k を $\min_i |k\boldsymbol{w}^T \boldsymbol{x}(i) - kb| = 1$ が成り立つように選ぶ．

さて，$k\boldsymbol{w}$ を改めて $\boldsymbol{w}$ と，また，kb を改めて b と定義する．すると，

$$\min_i \frac{|\boldsymbol{w}^T \boldsymbol{x}(i) - b|}{\|\boldsymbol{w}\|}$$

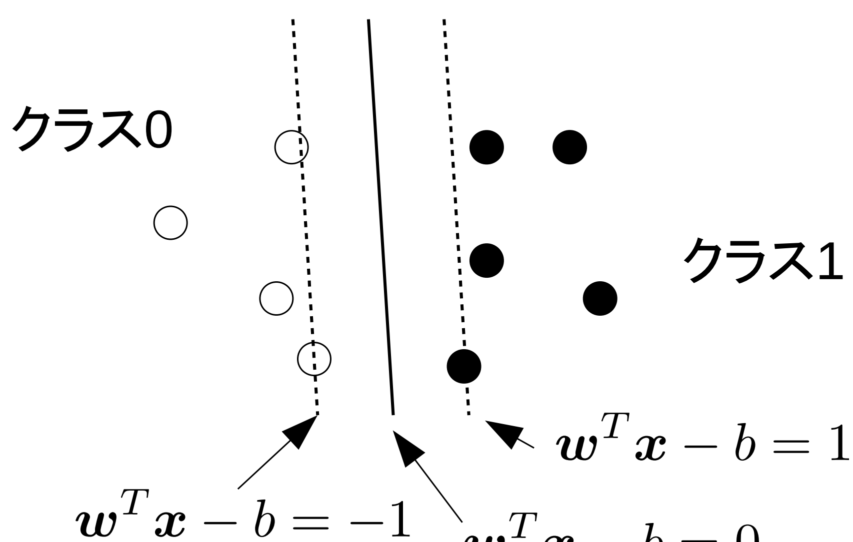
の最大化は $\|\boldsymbol{w}\|$ を最小にすることによって達成できる．

20. サポートベクターマシン

- 良い $\boldsymbol{w}$ と b を求める
  - 上記の議論をまとめると, 良い $\boldsymbol{w}$ と b は, 次の最適化問題を解くことによって得ることができる.

$$\text{Minimize } \frac{1}{2}\|\boldsymbol{w}\|^2 \text{ subject to } y(i)(\boldsymbol{w}^T\boldsymbol{x}(i) - b) \geq 1.$$

  - この最適化問題を解いて得られる結果は, サポートベクターマシンのハードマージンモデルと呼ばれる.



20. サポートベクターマシン

- 線形境界で正しく分類できないデータがある場合
  - 分類の条件を

    $$y(i)(\boldsymbol{w}^T\boldsymbol{x}(i)-b)\geq 1$$

    から

    $$y(i)(\boldsymbol{w}^T\boldsymbol{x}(i)-b)\geq 1-\varepsilon_i, \quad \varepsilon_i>0.$$

    に緩める.
  - ε_i の値が大きいと, 誤分類をより許容することになる. したがって, ε_i の値は小さい方が望ましい. そこで, 次の問題を定式化する.

    $$\text{Minimize } \frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_i\varepsilon_i,\ C>0$$

    $$\text{subject to } y(i)(\boldsymbol{w}^T\boldsymbol{x}(i)-b)\geq 1-\varepsilon,\ \varepsilon>0,$$

    これは, サポートベクターマシンのソフトマージンモデルを与える.

20. サポートベクターマシン

- データが線形境界ではうまく分類できない場合
  - データの次元が高いと, 線形境界によって分類するのは容易になる.
    データ個数よりも次元の方が値が大きければ, データは必ず線形境界で分類できる.
  - 上記を利用するために, 元のデータをより次元が高いベクトルに変換する非線形変換 $\phi(\boldsymbol{x})$ が使用される.
  - サポートベクターマシンの設計はデータベクトルの内積に基づいて行われる. もし, 非線形変換された後の内積が, 元のデータから直接計算できるなら, 言い換えると, もし次が成り立つ関数 K が存在するなら

    $$\phi(\boldsymbol{x}_1)^T\phi(\boldsymbol{x}_2) = K(\boldsymbol{x}_1, \boldsymbol{x}_2)$$

    設計は容易になる.

20. サポートベクターマシン

- 関数 K はカーネルと呼ばれる.
  - よく使われるカーネルは次のもの

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp\left(-\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2}{2\sigma^2}\right).$$