サポートベクタマシン
Support Vector Machine

§ SVM

SVM

= マージン最大化 + カーネルトリック

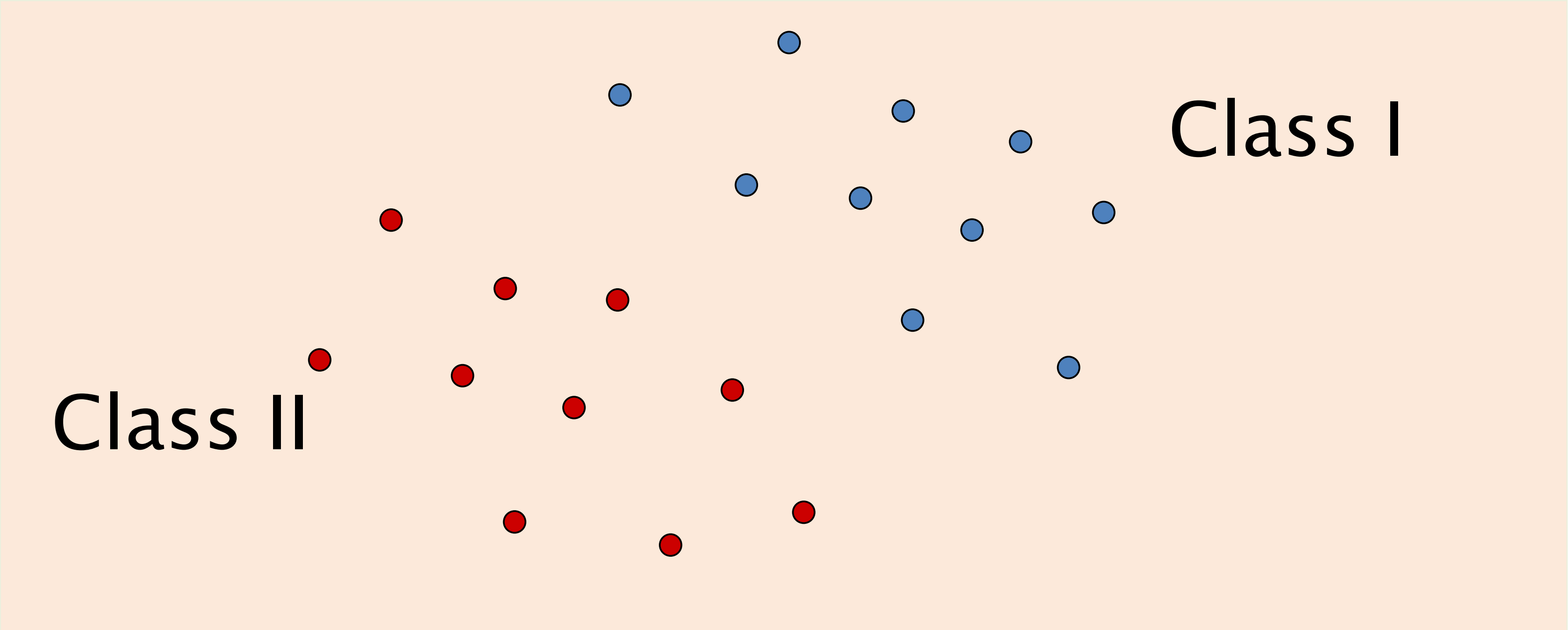
線形モデル

- 線形分離可能な問題の定式化
- 線形分離不可能な問題の定式化

非線形モデル

- 次元拡張
- カーネルトリック

§ マージン最大化



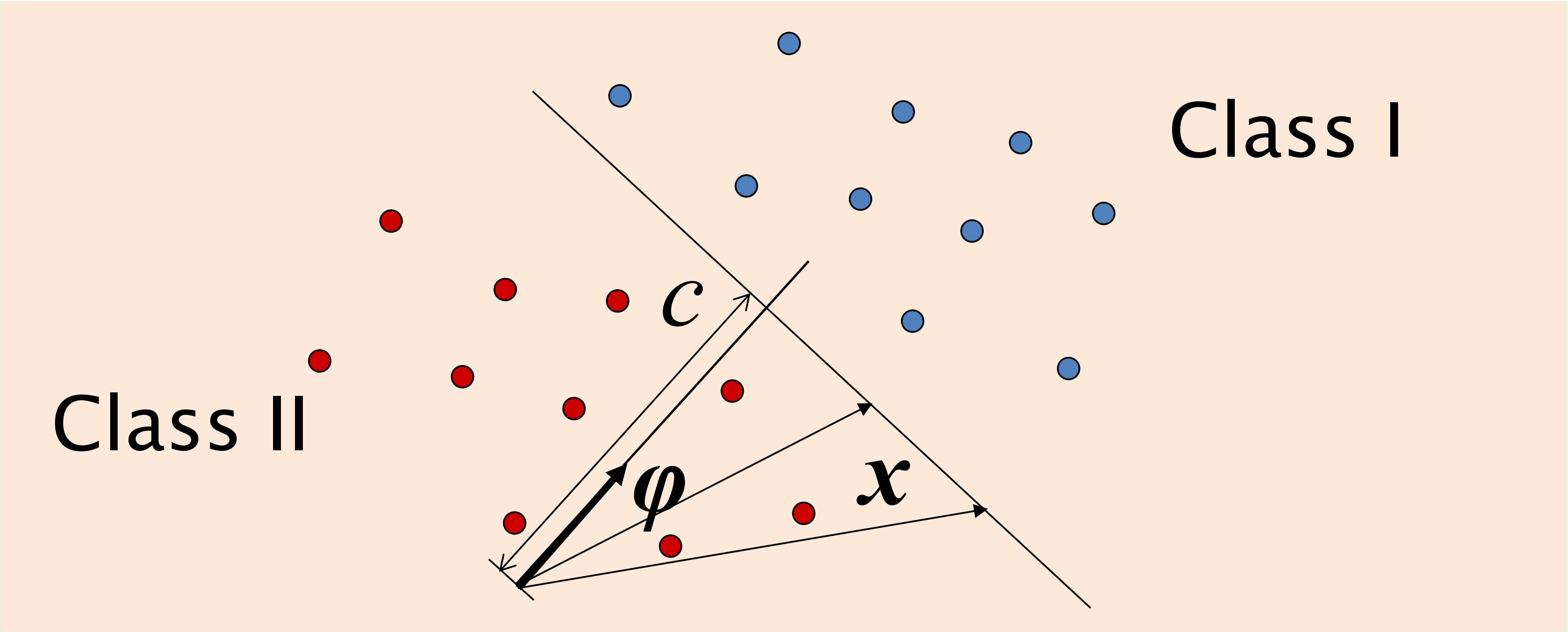
次の学習データが与えられたとする。

$$(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \cdots, (y_m, \boldsymbol{x}_m), \qquad \boldsymbol{x} \in R^n, y \in \{-1,1\}$$

ただし，

$$y = \begin{cases} 1 & \cdots \quad Class\ I \\ -1 & \cdots \quad Class\ II \end{cases}$$

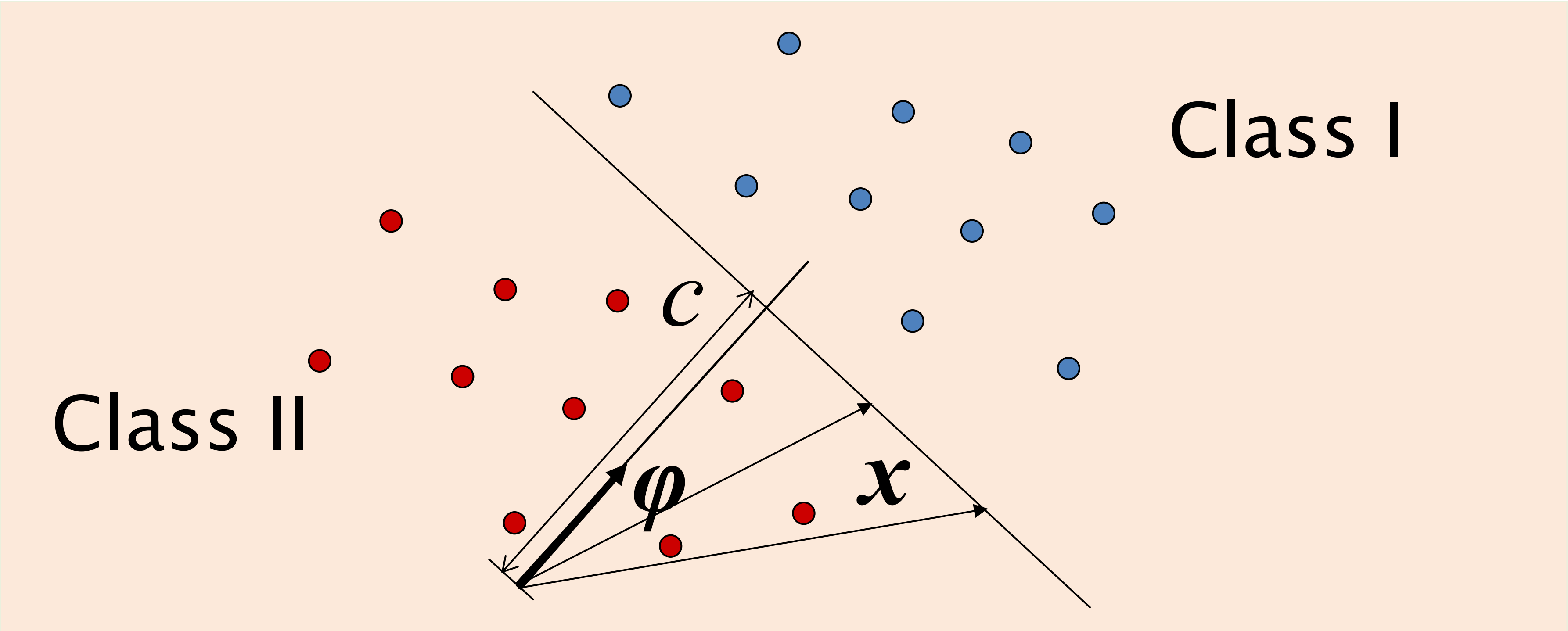
マージン最大化



カテゴリ I とカテゴリ II は，以下の超平面で分離可能であったとする。

$$\boldsymbol{\varphi}^T \cdot \boldsymbol{x} = c,\ \ \|\boldsymbol{\varphi}\| = 1$$

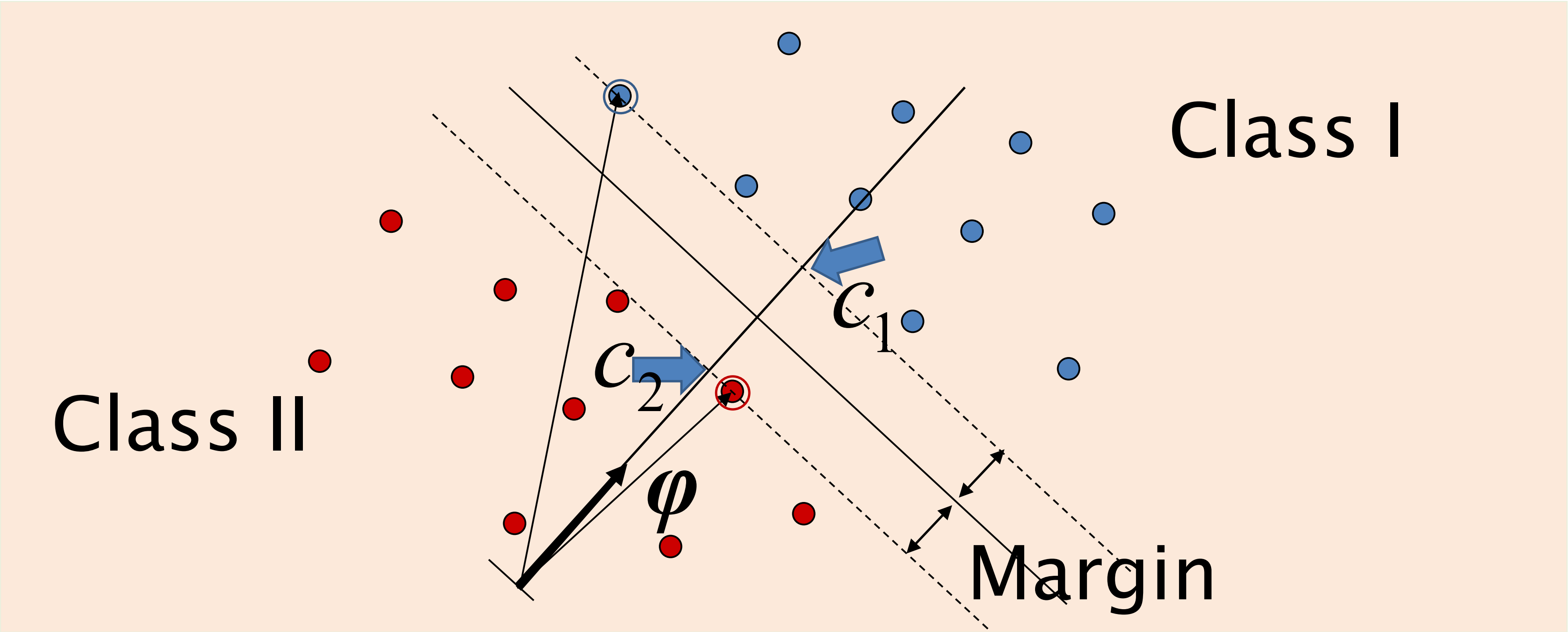
マージン最大化



ここで,

$$\boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i > c, \qquad if \ \ \boldsymbol{x}_i \in Class\ I$$
$$\boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i < c, \qquad if \ \ \boldsymbol{x}_i \in Class\ II$$

マージン最大化

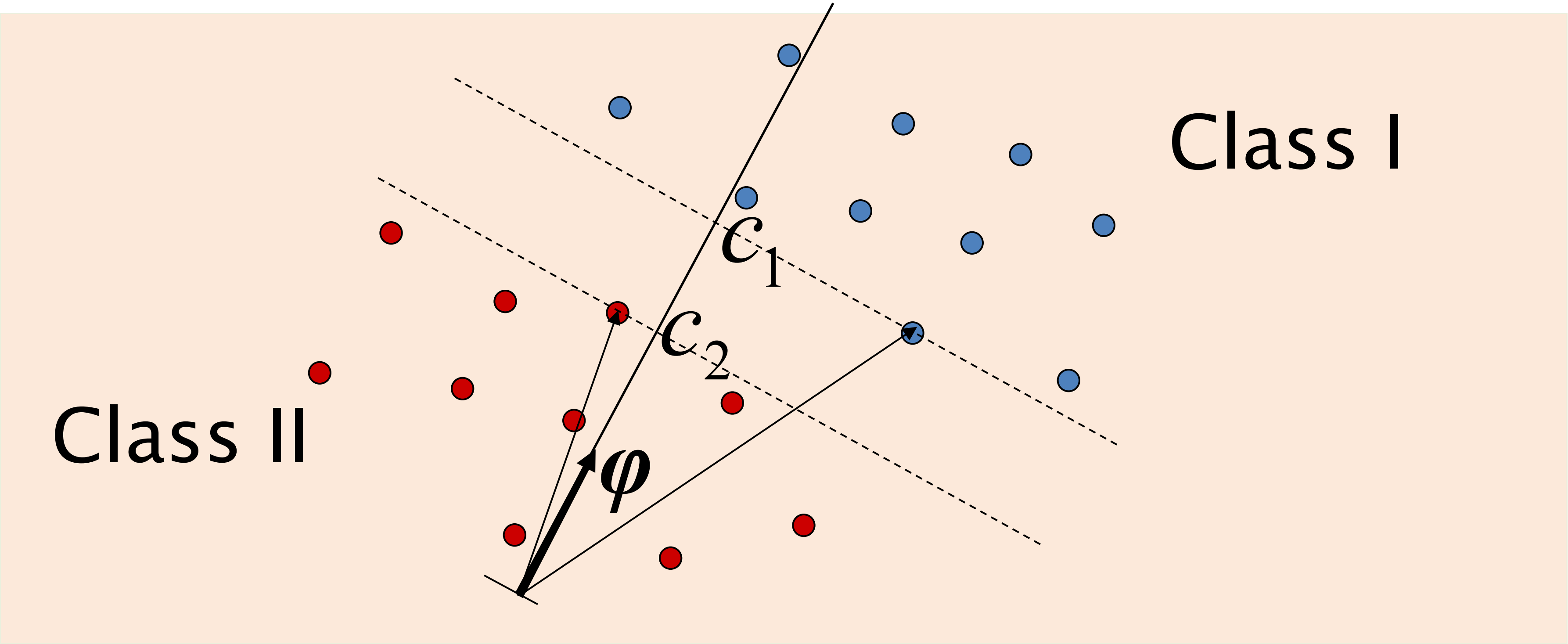


以下のようにc_1, c_2を定義する。

$$c_1(\boldsymbol{\varphi}) = \min_{\boldsymbol{x}_i \in I} \boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i\,, c_2(\boldsymbol{\varphi}) = \max_{\boldsymbol{x}_i \in II} \boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i\,.$$

$\rho = (c_1 - c_2)/2$ を“マージン”と呼ぶ．

マージン最大化

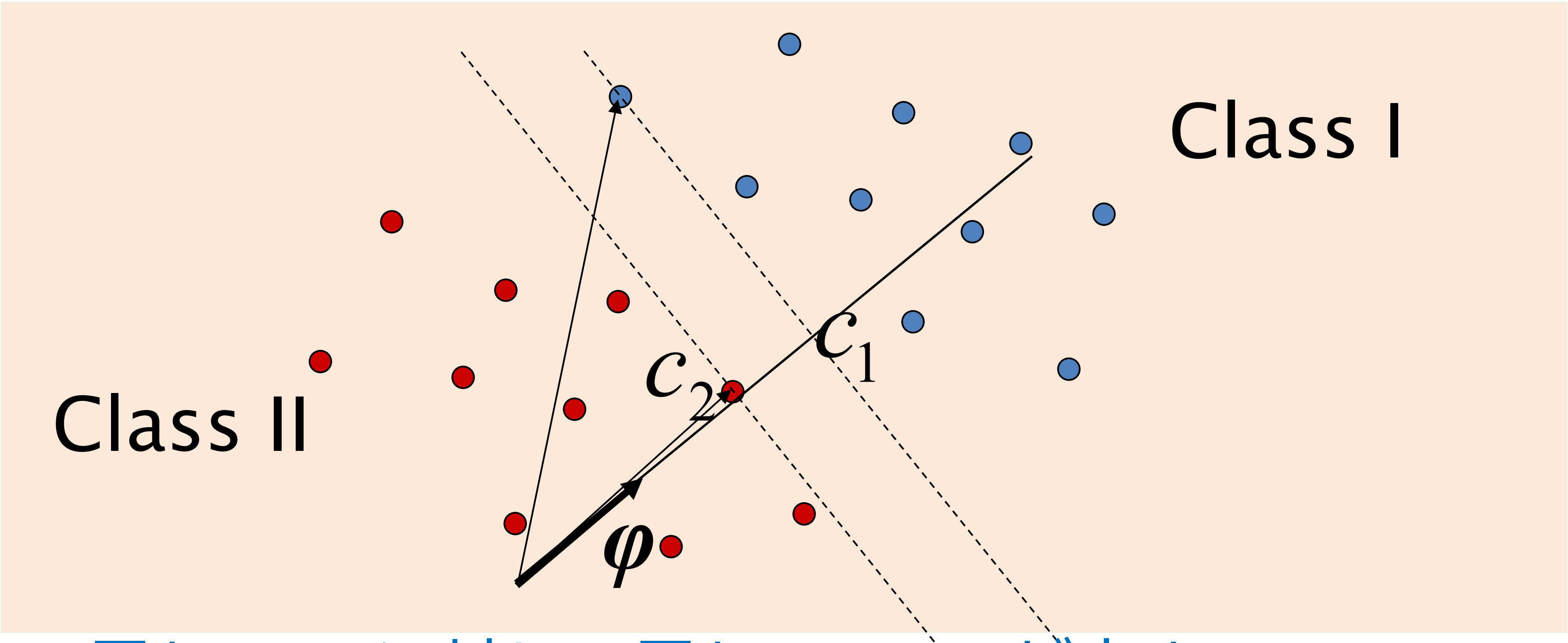


異なる $\boldsymbol{\varphi}$ に対し，異なる c_1 ,c_2が定まる。

$$c_1(\boldsymbol{\varphi}) = \min_{\boldsymbol{x}_i \in I} \boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i$$

$$c_2(\boldsymbol{\varphi}) = \max_{\boldsymbol{x}_i \in II} \boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i$$

マージン最大化

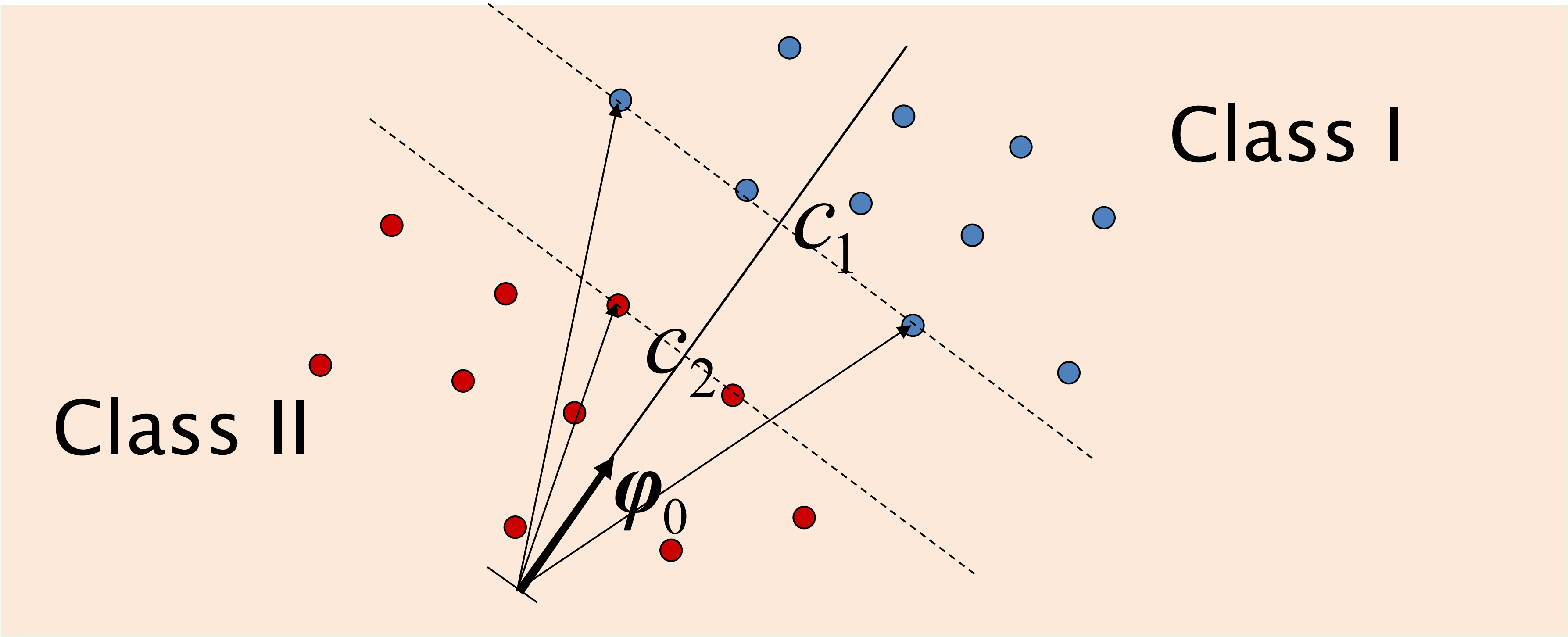


異なる $\boldsymbol{\varphi}$ に対し，異なる c_1 ,c_2が定まる。

$$c_1(\boldsymbol{\varphi}) = \min_{\boldsymbol{x}_i \in I} \boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i$$

$$c_2(\boldsymbol{\varphi}) = \max_{\boldsymbol{x}_i \in II} \boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i$$

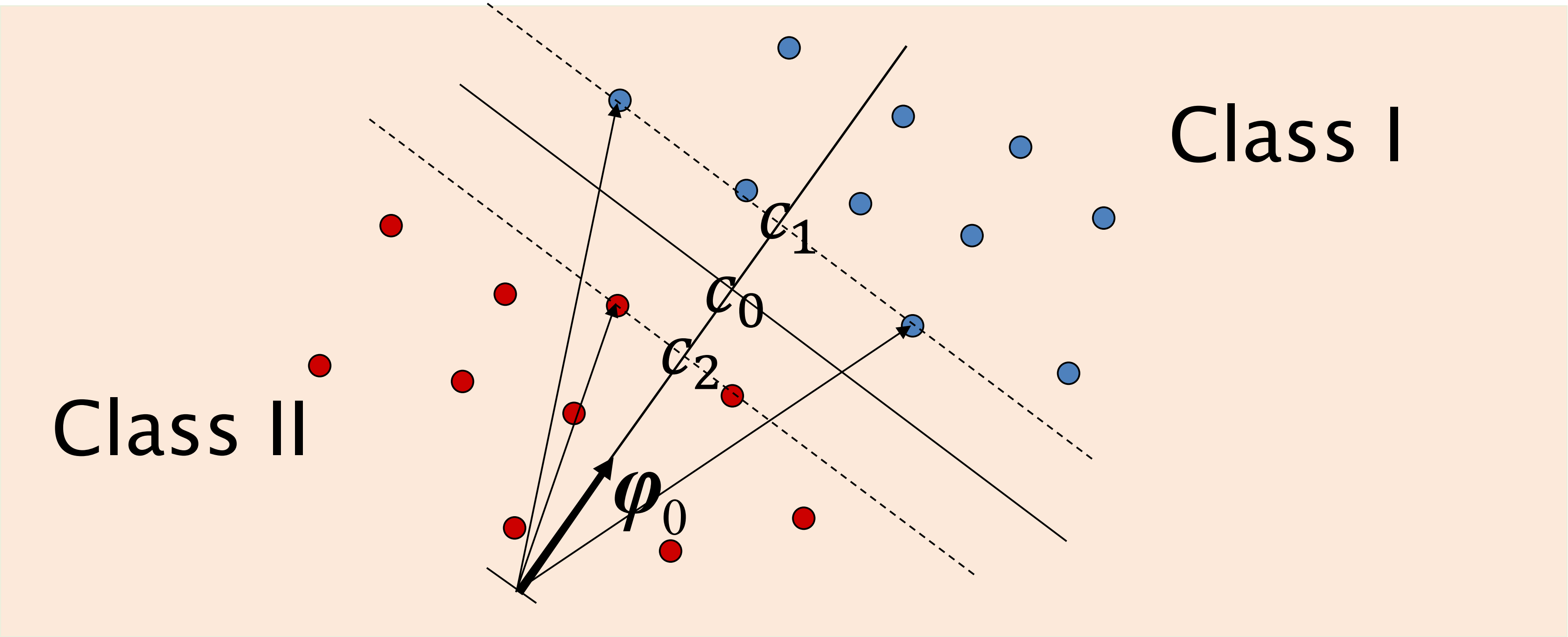
マージン最大化



マージンを最大化する φ_0 を考える。

$$\boldsymbol{\varphi}_0 = \underset{\boldsymbol{\varphi}}{\operatorname{argmax}}\, \rho(\boldsymbol{\varphi}) = \underset{\boldsymbol{\varphi}}{\operatorname{argmax}} \frac{c_1(\boldsymbol{\varphi}) - c_2(\boldsymbol{\varphi})}{2}$$

マージン最大化



このとき，次式は 最も危険の少ない識別境界となる。

$$\boldsymbol{\varphi}_0{}^T \cdot \boldsymbol{x} = c_0, \qquad \|\boldsymbol{\varphi}_0\| = 1$$

ここで，

$$c_0 = \frac{c_1(\boldsymbol{\varphi}_0) + c_2(\boldsymbol{\varphi}_0)}{2}$$

SVMの問題

$$\boldsymbol{\varphi}_0 = \underset{\boldsymbol{\varphi}}{\operatorname{argmax}}\, \rho(\boldsymbol{\varphi}) = \underset{\boldsymbol{\varphi}}{\operatorname{argmax}} \frac{c_1(\boldsymbol{\varphi}) - c_2(\boldsymbol{\varphi})}{2}. \quad (1)$$

s.t.

$$\begin{aligned} \boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i - c &\geq \rho, \quad && if \;\; y_i = 1 \\ \boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i - c &\leq -\rho, \quad && if \;\; y_i = -1 \end{aligned} \quad (2)$$

ここで,

$$\|\boldsymbol{\varphi}\| = 1,$$

$$c_1(\boldsymbol{\varphi}) = \min_{\boldsymbol{x}_i \in I} \boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i, \quad c_2(\boldsymbol{\varphi}) = \max_{\boldsymbol{x}_i \in II} \boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i\,, \quad (3)$$

$$c = \frac{c_1(\boldsymbol{\varphi}) + c_2(\boldsymbol{\varphi})}{2}$$

等価な問題

$$\max_{\boldsymbol{\varphi}} \rho(\boldsymbol{\varphi}) = \max_{\boldsymbol{\varphi}} \frac{c_1(\boldsymbol{\varphi}) - c_2(\boldsymbol{\varphi})}{2}$$

$$\|\boldsymbol{\varphi}\| = 1$$

$s.t.$

$$\boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i - c \geq \rho, \quad if \;\; y_i = 1$$
$$\boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i - c \leq -\rho, \quad if \;\; y_i = -1$$

$$c = \frac{c_1(\boldsymbol{\varphi}) + c_2(\boldsymbol{\varphi})}{2}$$

$\Leftrightarrow$

$$\max_{\boldsymbol{\psi}} \rho\left(\frac{\boldsymbol{\psi}}{\|\boldsymbol{\psi}\|}\right) = \max_{\boldsymbol{\psi}} \frac{\rho'(\boldsymbol{\psi})}{\|\boldsymbol{\psi}\|}$$

$s.t.$

$$\boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b \geq \rho'(\boldsymbol{\psi}), \quad if \;\; y_i = 1$$
$$\boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b \leq -\rho'(\boldsymbol{\psi}), \quad if \;\; y_i = -1$$

$$\leftarrow \quad \boldsymbol{\varphi} = \frac{\boldsymbol{\psi}}{\|\boldsymbol{\psi}\|},$$
$$\rho\left(\frac{\boldsymbol{\psi}}{\|\boldsymbol{\psi}\|}\right) = \frac{\rho'(\boldsymbol{\psi})}{\|\boldsymbol{\psi}\|},$$
$$b = c \cdot \|\boldsymbol{\psi}\|$$

$$\max_{\boldsymbol{\psi}} \rho\left(\frac{\boldsymbol{\psi}}{\|\boldsymbol{\psi}\|}\right) = \max_{\boldsymbol{\psi}} \frac{\rho'(\boldsymbol{\psi})}{\|\boldsymbol{\psi}\|}$$

$s.t.$

$$\boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b \geq \rho'(\boldsymbol{\psi}), \quad if \;\; y_i = 1$$
$$\boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b \leq -\rho'(\boldsymbol{\psi}), \quad if \;\; y_i = -1$$

$$\leftarrow \max_{\boldsymbol{\psi}} \frac{\rho'(\boldsymbol{\psi})}{\|\boldsymbol{\psi}\|}$$
$$\Leftrightarrow \min_{\boldsymbol{\psi}} \|\boldsymbol{\psi}\|, \rho'(\boldsymbol{\psi}) = 1$$

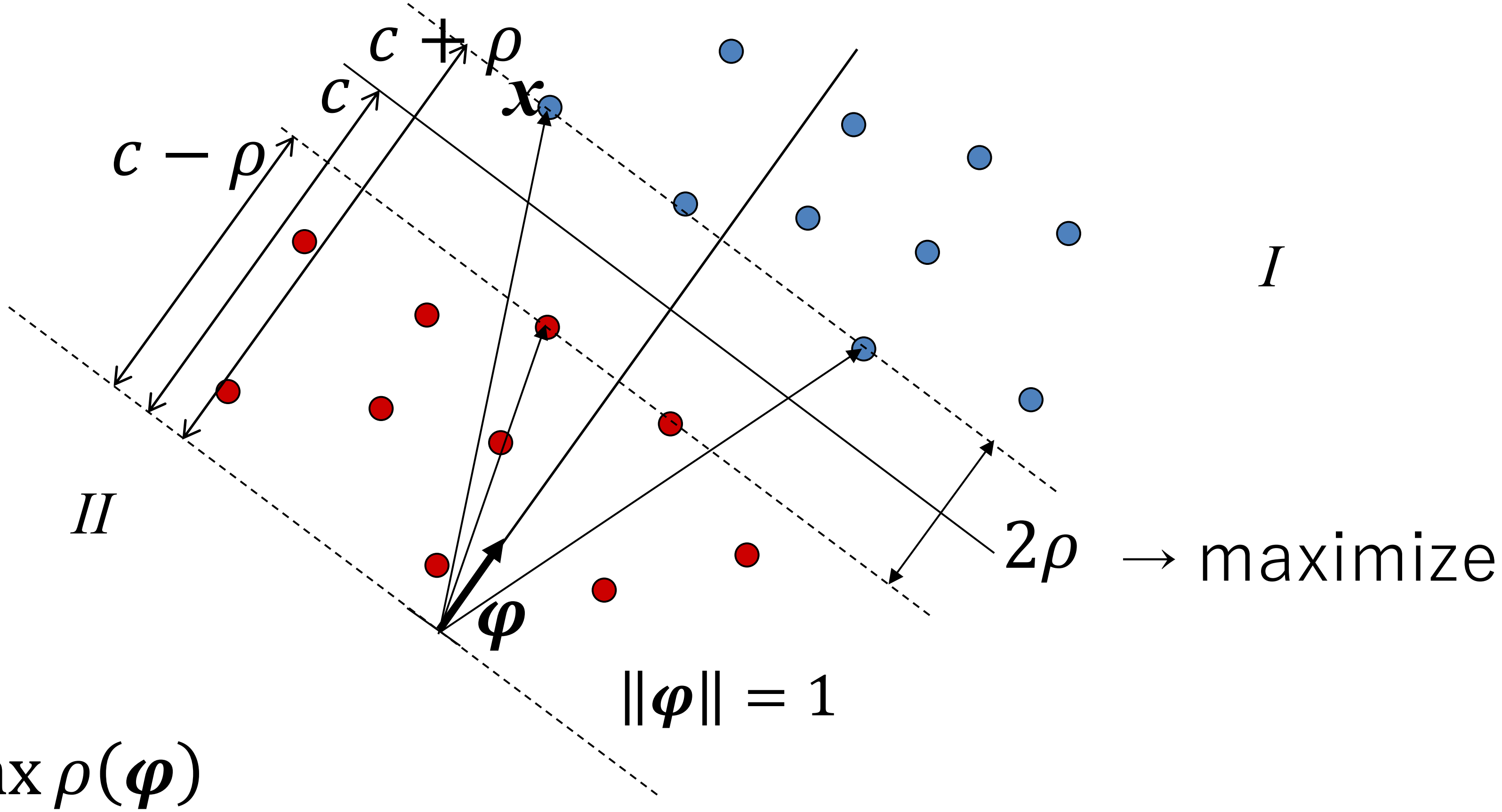
$\Leftrightarrow$

$$\min_{\boldsymbol{\psi}} \|\boldsymbol{\psi}\| \tag{4}$$

$s.t.$

$$\boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b > 1, \quad if \;\; y_i = 1$$
$$\boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b < -1, \quad if \;\; y_i = -1 \tag{5}$$

マージン最大化



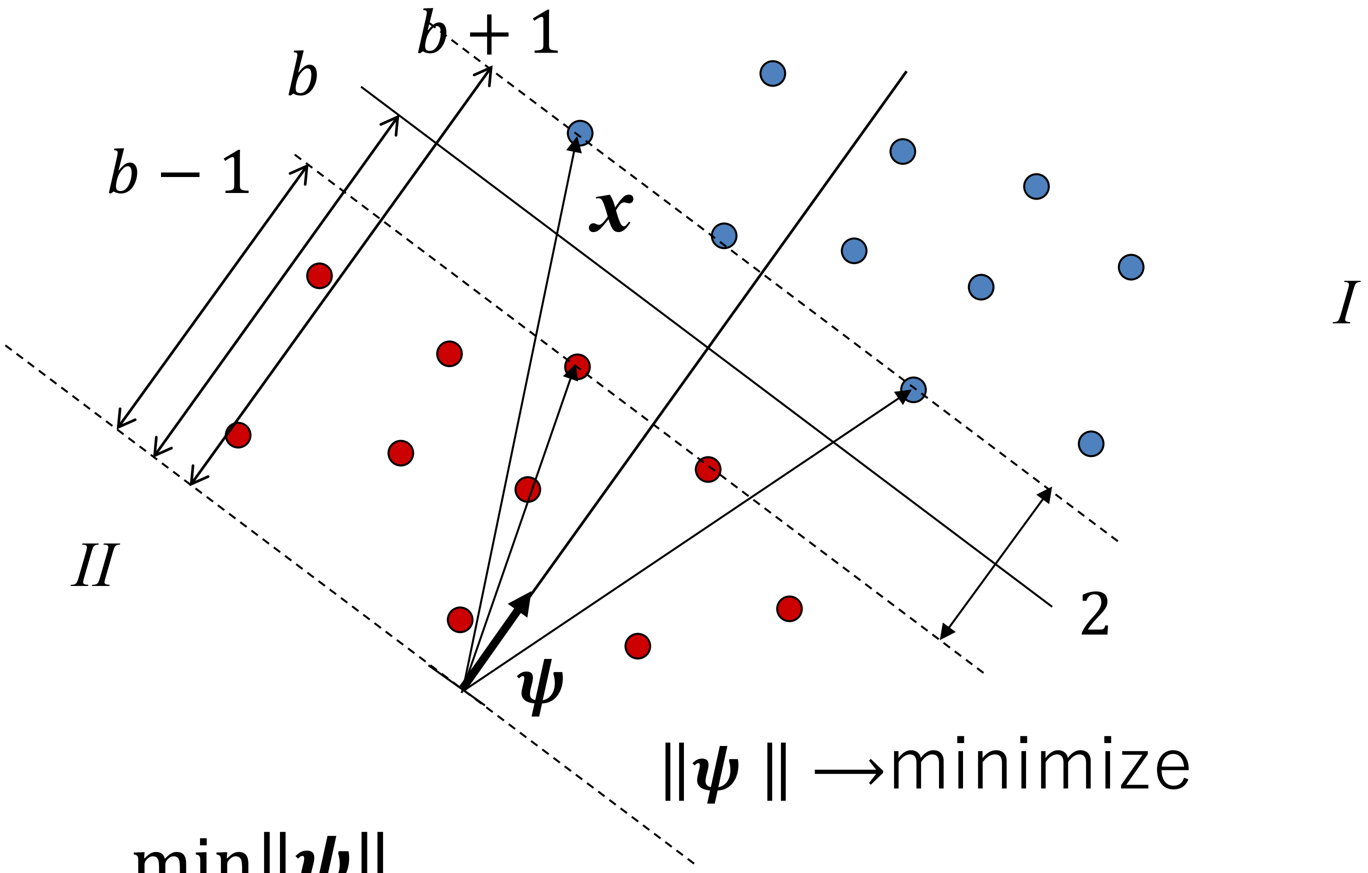
$$\max_{\boldsymbol{\varphi}} \rho(\boldsymbol{\varphi})$$

$s.t.$

$$\boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i - c \geq \rho(\boldsymbol{\varphi}), \quad if\ \ y_i = 1$$
$$\boldsymbol{\varphi}^T \cdot \boldsymbol{x}_i - c \leq -\rho(\boldsymbol{\varphi}), \quad if\ \ y_i = -1$$

$$\|\boldsymbol{\varphi}\| = 1$$

マージン最大化



$$\min_{\boldsymbol{\psi}} \|\boldsymbol{\psi}\|$$

$s.t.$

$$\boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b \geq 1, \qquad if\ \ y_i = 1$$
$$\boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b \leq -1, \qquad if\ \ y_i = -1$$

SVM の問題

$$\boldsymbol{\psi_0} = \underset{\boldsymbol{\psi}}{\operatorname{argmin}} \|\boldsymbol{\psi}\| \qquad (4)$$

s.t.

$$\begin{aligned} \boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b > 1, \quad & if \;\; y_i = 1 \\ \boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b < -1, \quad & if \;\; y_i = -1 \end{aligned} \qquad (5)$$

または,

$$\boldsymbol{\psi_0} = \underset{\boldsymbol{\psi}}{\operatorname{argmin}} \frac{1}{2} \boldsymbol{\psi}^T \boldsymbol{\psi} \qquad (6)$$

s.t.

$$\begin{aligned} \boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b > 1, \quad & if \;\; y_i = 1 \\ \boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b < -1, \quad & if \;\; y_i = -1 \end{aligned} \qquad (5)$$

ラグランジュ定数$\boldsymbol{\alpha}$を導入してラグランジュ関数を定める

$$
\begin{aligned}
L(\boldsymbol{\psi}, b, \boldsymbol{\alpha}) &= \overbrace{\frac{1}{2}\boldsymbol{\psi}^T\boldsymbol{\psi}}^{\text{目的関数}} - \sum_{i=1}^{m} \overbrace{\alpha_i}^{\text{ラグランジュ定数}} \overbrace{\left(y_i(\boldsymbol{\psi}^T\boldsymbol{x}_i - b) - 1\right)}^{\text{制約}} \\
&= \frac{1}{2}\boldsymbol{\psi}^T\boldsymbol{\psi} - \boldsymbol{\alpha}^T D^T\boldsymbol{\psi} + b\boldsymbol{\alpha}^T\boldsymbol{y} + \|\boldsymbol{\alpha}\|_1 \qquad (7)
\end{aligned}
$$

$$D = (y_1\boldsymbol{x}_1, y_2\boldsymbol{x}_2, \cdots, y_m\boldsymbol{x}_m)$$

双対問題を導くために，設計変数 $\boldsymbol{\psi}, b$ で，$L(\boldsymbol{\psi}, b, \boldsymbol{\alpha})$ を偏微分

$$\frac{\partial}{\partial\boldsymbol{\psi}}L(\boldsymbol{\psi}, b, \boldsymbol{\alpha}) = \boldsymbol{\psi} - D\boldsymbol{\alpha} = 0$$

$$\frac{\partial}{\partial b}L(\boldsymbol{\psi}, b, \boldsymbol{\alpha}) = \boldsymbol{\alpha}^T\boldsymbol{y} = 0$$

以下の解を得る。

$$\boldsymbol{\psi} = D\boldsymbol{\alpha}, \quad \boldsymbol{\alpha}^T \boldsymbol{y} = 0 \tag{8}$$

これらを $L(\boldsymbol{\psi}, b, \boldsymbol{\alpha})$ に代入

$$\begin{aligned} L(\boldsymbol{\alpha}) &= \frac{1}{2}\boldsymbol{\alpha}^T D^T D\boldsymbol{\alpha} - \boldsymbol{\alpha}^T D^T D\boldsymbol{\alpha} + b \cdot \boldsymbol{\alpha}^T \boldsymbol{y} + \|\boldsymbol{\alpha}\|_1 \\ &= -\frac{1}{2}\boldsymbol{\alpha}^T D^T D\boldsymbol{\alpha} + \|\boldsymbol{\alpha}\|_1 \\ &= -\frac{1}{2}\boldsymbol{\alpha}^T A\boldsymbol{\alpha} + \|\boldsymbol{\alpha}\|_1 \end{aligned} \tag{9}$$

$$A = \left((y_i\boldsymbol{x}_i)^T(y_j\boldsymbol{x}_j)\right) = \left(y_i y_j \boldsymbol{x}_i{}^T \boldsymbol{x}_j\right)$$

双対問題は以下のとおりとなる。

$$\boldsymbol{\alpha}^0 = \max_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}) = \max_{\boldsymbol{\alpha}} \left(-\frac{1}{2} \boldsymbol{\alpha}^T A \boldsymbol{\alpha} + \|\boldsymbol{\alpha}\|_1 \right) \qquad (10)$$

$$A = \left(y_i y_j \boldsymbol{x}_i{}^T \boldsymbol{x}_j \right)$$

s.t.

$$\alpha_i \geq 0, \qquad i = 1,2,\cdots,m \qquad (11)$$

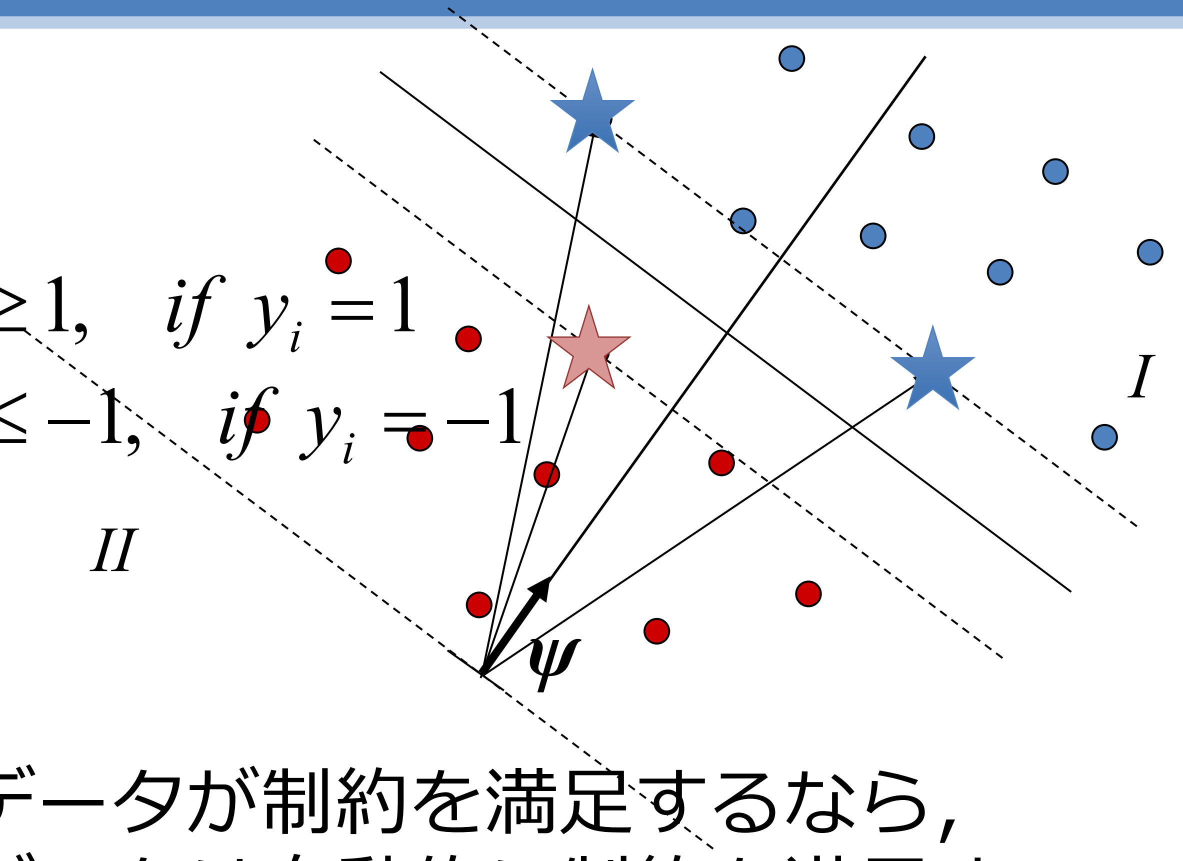
$\boldsymbol{\alpha}^0$, を用いて，最適な超平面の法線ベクトルは，以下のように表現できる。

$$\boldsymbol{\psi}^0 = D\boldsymbol{\alpha}^0 = \sum_{i=0}^{m} \alpha_i{}^0 y_i \boldsymbol{x}_i \qquad (12)$$

ここで， $\boldsymbol{\alpha}^0$ は， m (データ数)-次元のベクトル。

しかし，ほとんどの制約が非アクティブであるから，$\boldsymbol{\alpha}^0$ のほとんどの要素は 0

$$(\boldsymbol{x}_i * \boldsymbol{\psi}) - b \geq 1, \quad if\ \ y_i = 1$$
$$(\boldsymbol{x}_i * \boldsymbol{\psi}) - b \leq -1, \quad if\ \ y_i = -1$$



★★のデータが制約を満足するなら,
●● のデータは自動的に制約を満足する。

すなわち, ●● に対する制約は 非アクティブ である。

制約: $\boldsymbol{x}_i{}^T\boldsymbol{\psi} - b \geq 1, \qquad if \quad y_i = 1,$

$\boldsymbol{x}_i{}^T\boldsymbol{\psi} - b \leq -1, \qquad if \quad y_i = -1,$

サポートベクトル

Kuhn-Tuckerの関係式により,

$$\alpha_i{}^0(y_i(\boldsymbol{x}_i{}^T\boldsymbol{\psi}^0 \ - b^0) - 1) = 0$$

零でない $\alpha_i{}^0$ に対応する i の集合を Sとすると,

$$y_i(\boldsymbol{x}_i{}^T\boldsymbol{\psi}^0 \ - b^0) - 1 = 0, \qquad\qquad if \ \ i \in S$$

$\boldsymbol{x}_i$, $i \in S$ をサポートベクトルと呼ぶ。

サポートベクトル

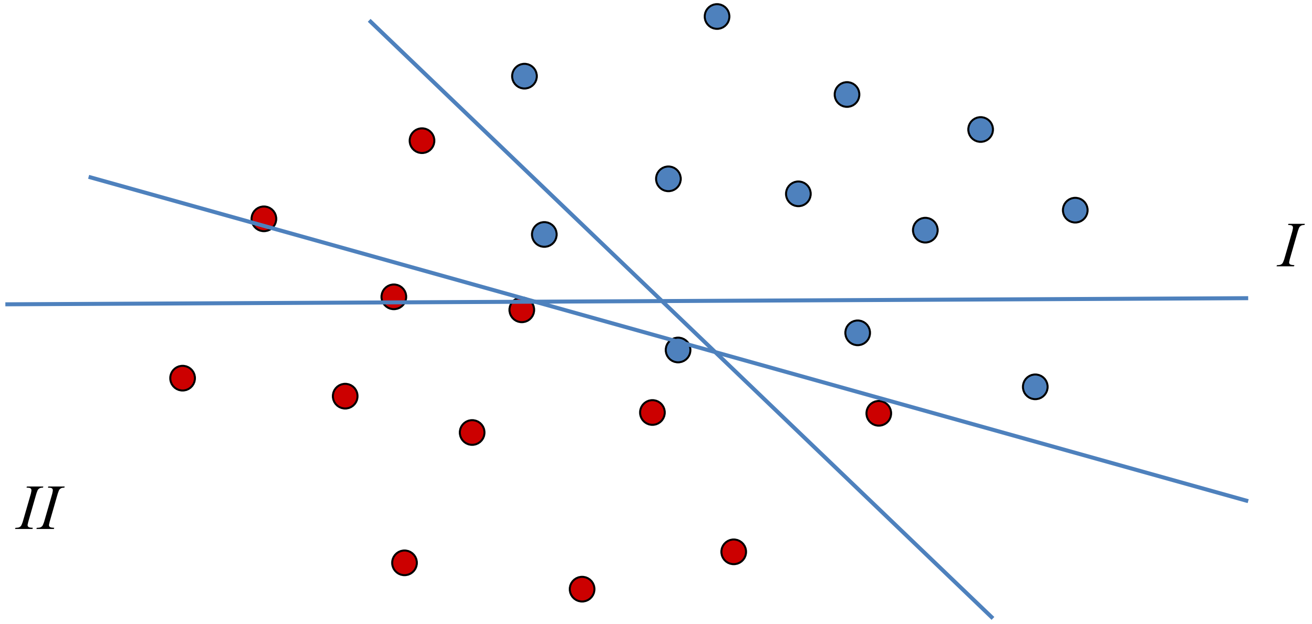
最適超平面は，サポートベクトルだけで決めることができる。

$$\boldsymbol{\psi}^0 = \sum_{i \in S} \alpha_i^0 y_i \boldsymbol{x}_i \qquad (13)$$

$$b^0 = \boldsymbol{x}_i{}^T \boldsymbol{\psi}^0 - y_i, \qquad i \in S$$

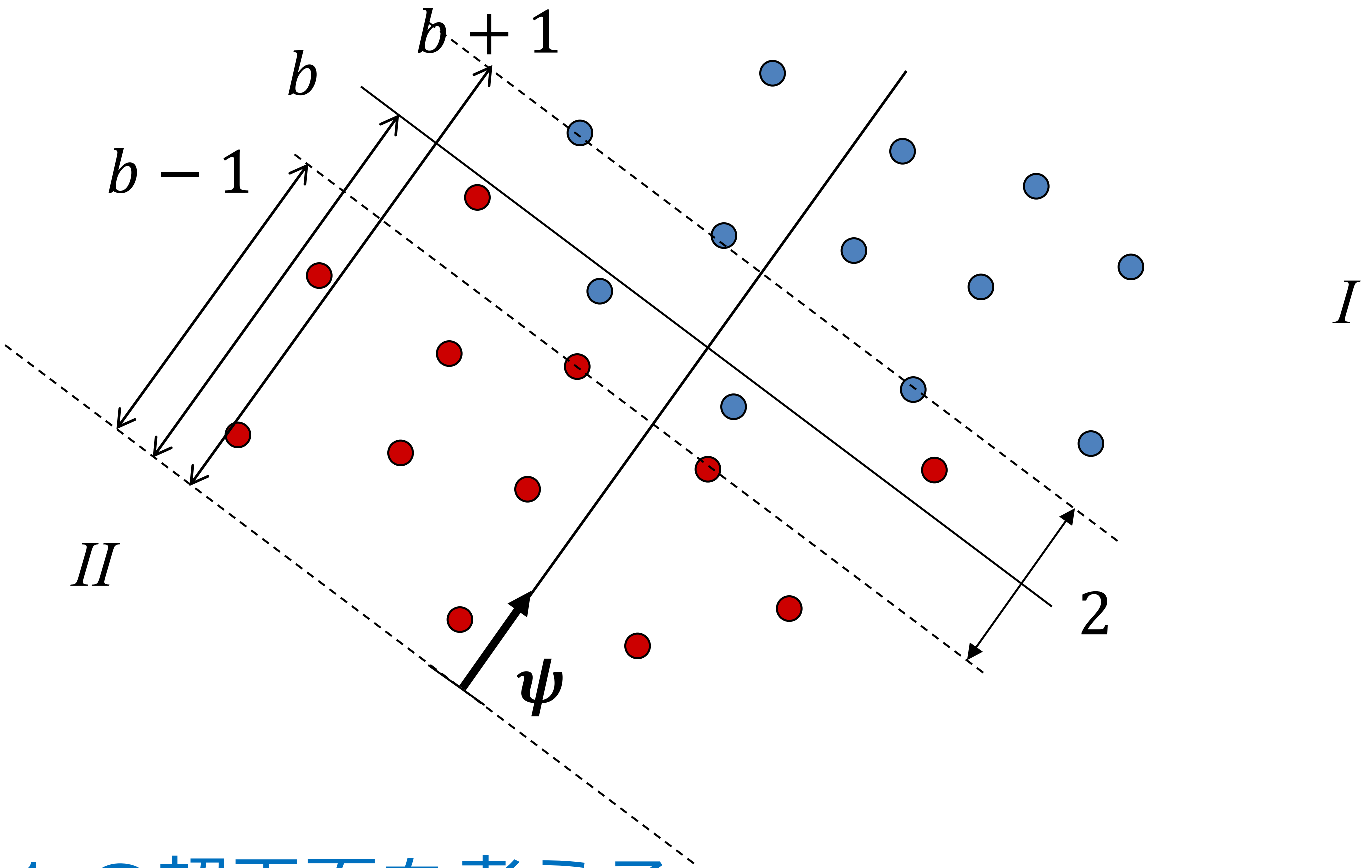
$$f(\boldsymbol{x}) = \boldsymbol{\psi}^0 \boldsymbol{x} = \sum_{i \in S} \left(\alpha_i^0 y_i \boldsymbol{x}_i{}^T \boldsymbol{x} \right) = b^0 \qquad (14)$$

§ 線形分離不可能な場合



線形分離不可能
= 与えられた複数のクラスのデータを
　分離する超平面が存在しない

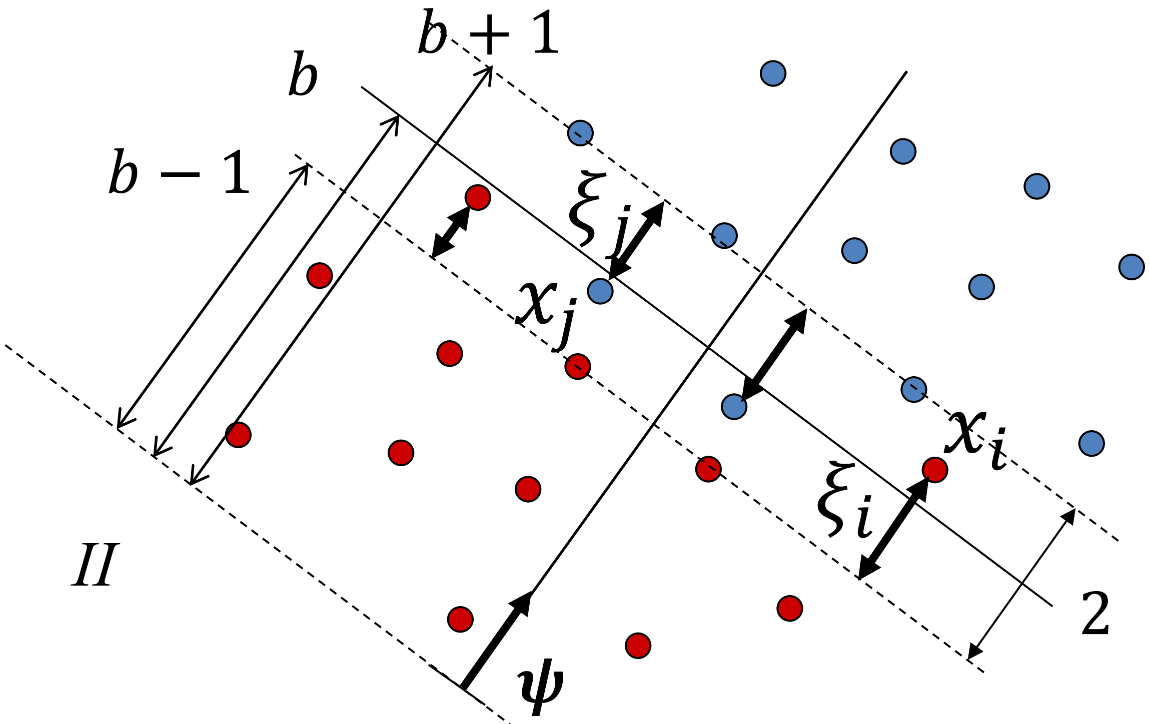
線形分離不可能な場合



マージン 1 の超平面を考える。

$$\boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b = 1,$$
$$\boldsymbol{\psi}^T \cdot \boldsymbol{x}_i - b = -1,$$

線形分離不可能な場合



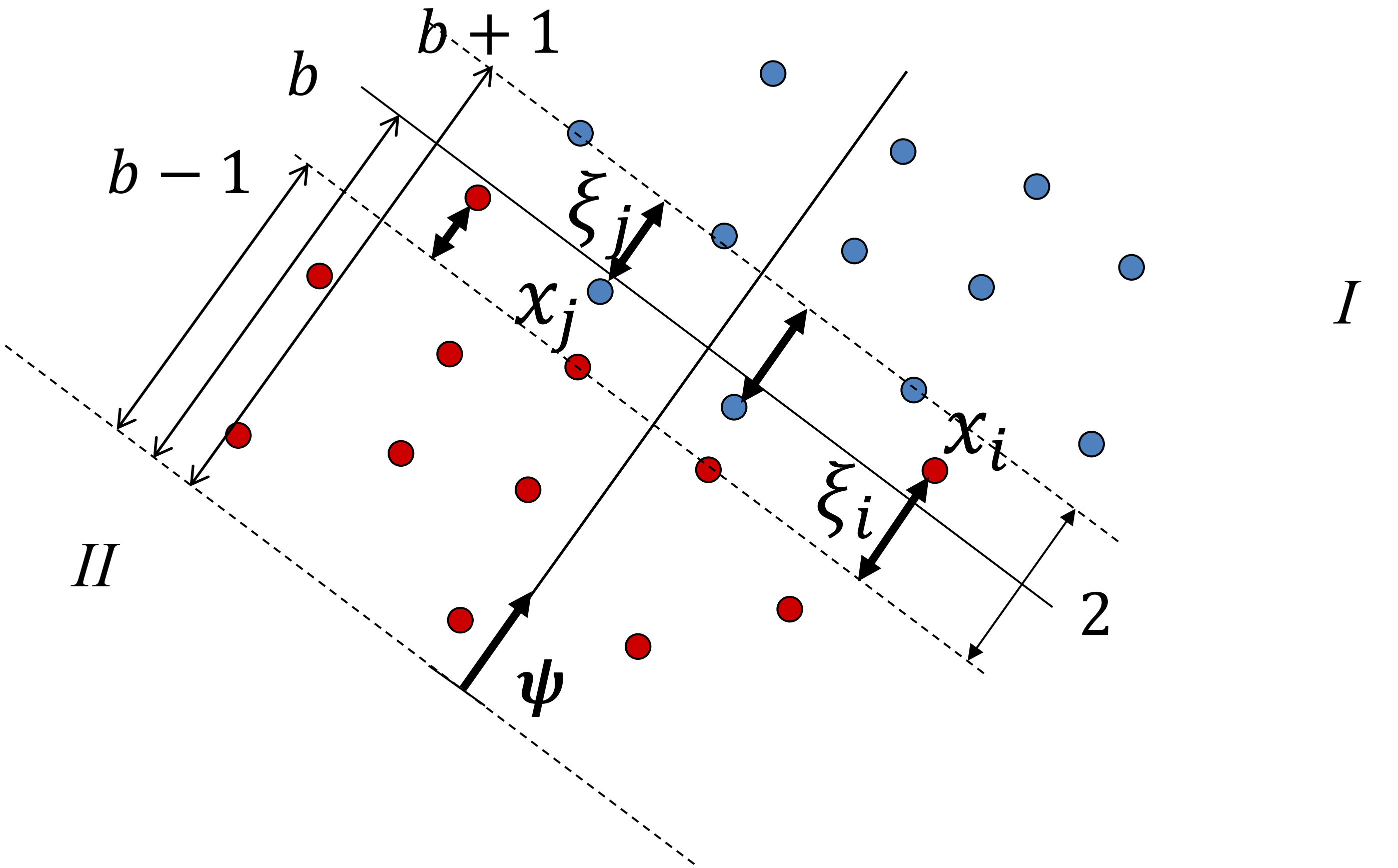
以下のロス関数を定義する。

$$F(\boldsymbol{\xi}) = \sum_{i=1}^{m} \xi_i , \qquad \xi_i \geq 0. \qquad i = 1,2,\cdots,m$$

ξ_i は，マージン１の超平面から相手クラス側にはみ出した距離。

$$y_i(\boldsymbol{\psi}^T \boldsymbol{x}_i - b) \geq 1 - \xi_i,$$

線形分離不可能な場合



マージンを大きくするために，$\|\boldsymbol{\psi}\|^2$ を小さくし，

エラーを小さくするために， $F(\boldsymbol{\xi}) = \sum_{i=1}^{m} \xi_i$, $\xi_i \geq 0$, を小さくする

線形分離不可能な場合

問題は,

$$\min\left(\frac{1}{2}\boldsymbol{\psi}^T\boldsymbol{\psi} + C\cdot\sum_{i=1}^{m}\xi_i\right) \tag{14}$$

$$s.t.\ \ y_i(\boldsymbol{\psi}^T\boldsymbol{x}_i - b) \geq 1 - \xi_i,\ \xi_i \geq 0,\ \ i = 1,2,\cdots,m$$

ラグランジュ関数は，定数 α, β を用いて以下のようになる。

$$L(\boldsymbol{\psi},\boldsymbol{\xi},b,\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{\psi}^T\boldsymbol{\psi} + C\cdot\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i(y_i(\boldsymbol{\psi}^T\boldsymbol{x}_i - b) - 1 + \xi_i) - \sum_{i=1}^{m}\beta_i\xi_i \tag{15}$$

線形分離不可能な場合

変形して,

$$L(\boldsymbol{\psi}, \boldsymbol{\xi}, b, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{\psi}^T \boldsymbol{\psi} + C \cdot \|\boldsymbol{\xi}\|_1 - \boldsymbol{\alpha}^T D\boldsymbol{\psi} + b\boldsymbol{\alpha}^T \boldsymbol{y} + \|\boldsymbol{\alpha}\|_1 - \boldsymbol{\alpha}^T \boldsymbol{\xi} - \boldsymbol{\beta}^T \boldsymbol{\xi} \tag{15'}$$

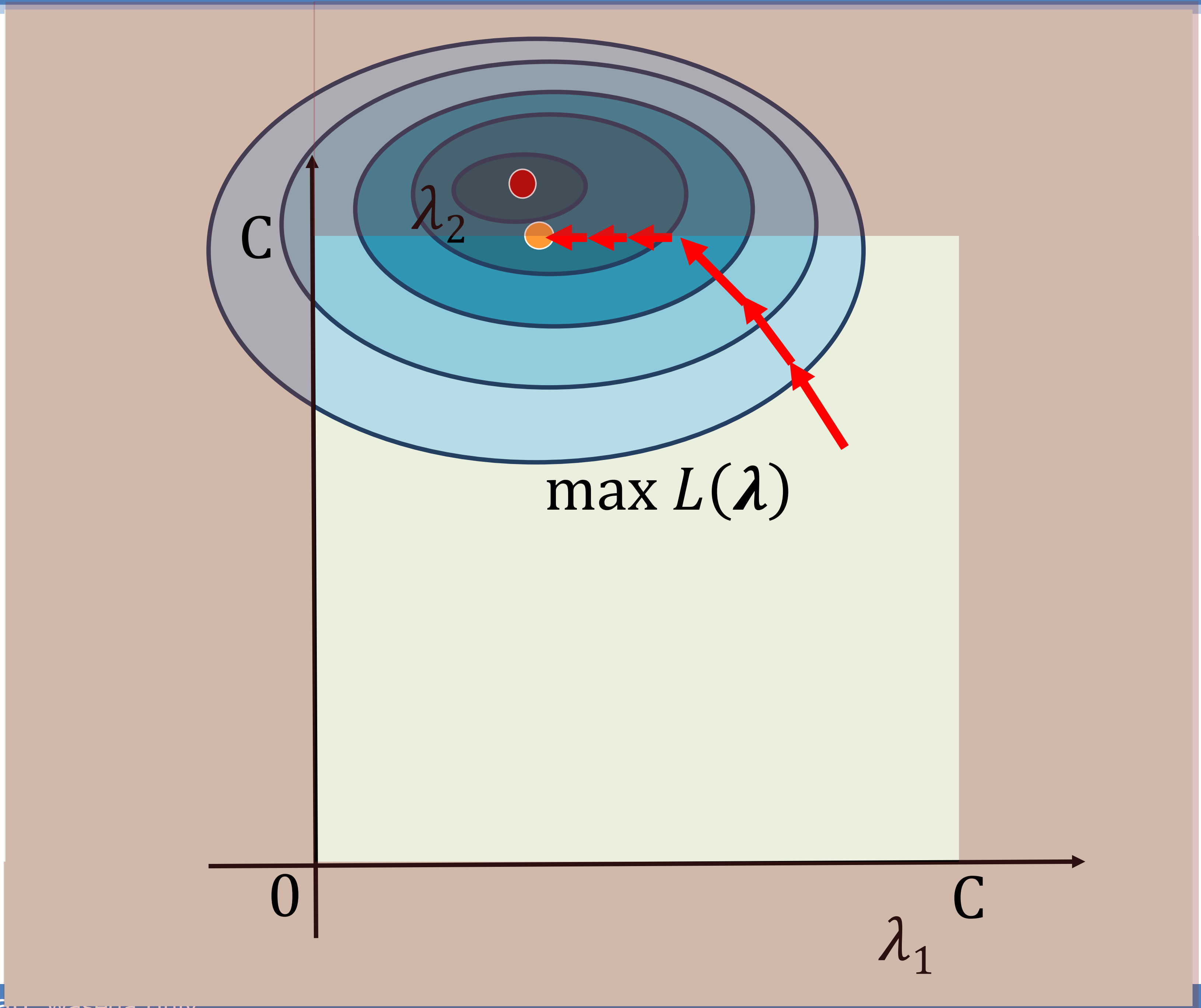
双対問題は,

$$\boldsymbol{\alpha}^0 = \max_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}) = \max_{\boldsymbol{\alpha}} \left(-\frac{1}{2}\boldsymbol{\alpha}^T A\boldsymbol{\alpha} + \|\boldsymbol{\alpha}\|_1\right)$$

$$A = \left(y_i y_j \boldsymbol{x}_i{}^T \boldsymbol{x}_j\right) \tag{16}$$

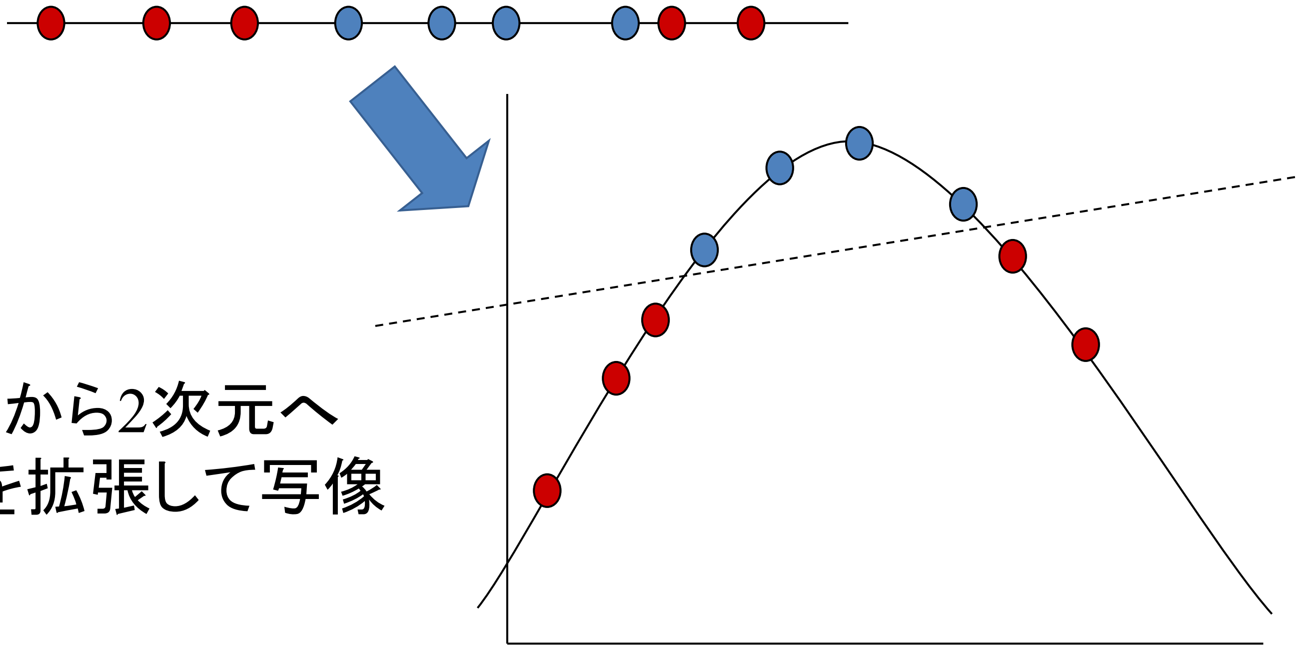
s.t.

$$0 \le \alpha_i \le C, \qquad i = 1,2,\cdots,m \tag{17}$$


C
λ_2
max $L(\boldsymbol{\lambda})$
0
C
λ_1

§ 次元拡張

- 線形識別不可能なデータも，次元を拡張すると線形識別可能になることがある。



1次元から2次元へ
次元を拡張して写像

- カーネルトリック：ラグランジュ関数の中の $\boldsymbol{x}_i{}^T\boldsymbol{x}_j$ を $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ で置き換える。$\boldsymbol{x}_i$ を高次元のベクトル $\boldsymbol{\phi}(\boldsymbol{x}_i)$ に写像してマージン最大化するのと同じ効果を持つ。

まとめ

- マージンを最大化する基準で識別境界を決める手法をサポートベクタマシン（SVM）と呼ぶ。
- SVMの設計は，不等式制約の最適化問題を解くことで実現できる。

演習問題

1. SVMの説明にある，式(14) を導け。

2. サポートベクタに対するラグランジュ係数を$\boldsymbol{\alpha}^0$，マージンをρとするとき，

$$\boldsymbol{\psi}^{0^T}\boldsymbol{\psi}^0 = \sum_{i=1}^{m} \alpha_i^0$$

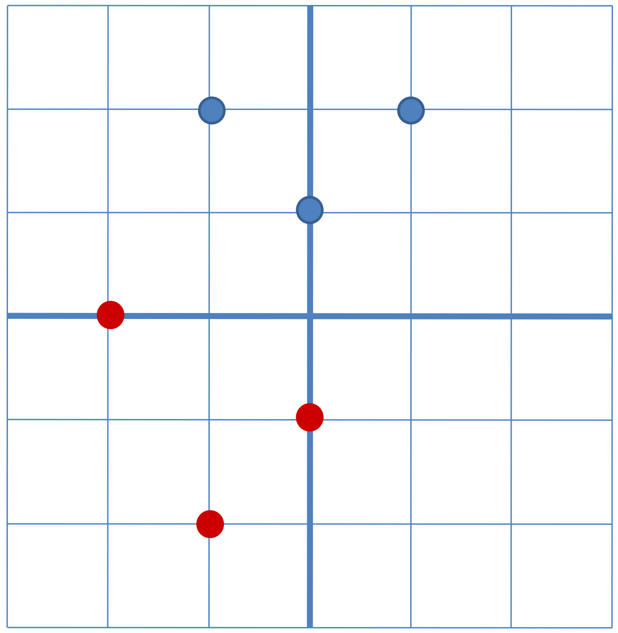
$$\rho(\boldsymbol{\psi}^0) = \frac{1}{\|\boldsymbol{\psi}^0\|}$$

$$L(\boldsymbol{\alpha}) < L(\boldsymbol{\alpha}^0) = \frac{1}{2}\frac{1}{\rho(\boldsymbol{\psi}^0)^2}, \quad \boldsymbol{\alpha} \neq \boldsymbol{\alpha}^0$$

を証明せよ。

演習問題

3. カテゴリAとして，(1,2),(0,1),(-1,2)が， カテゴリBとして，(0,-1),(-1,-2),(-2,0) が与えられたとき，これらのカテゴリをマージン最大化基準で分離する直線を求めよ。



演習問題

4. SVMの説明にある，式(17) を導け。

5. クラス A のデータが，(1,2),(0,1),(-1,0),(-1,2)，クラス B のデータが (0,-1),(-1,-2),(-1,1),(-2,0) と与えられている。このとき，マージン最大化の基準で識別器を設計せよ。