演習　カーネル法

瀬戸 道生（防衛大学校・数学教育室）

自己紹介など

１５年前：横浜

神奈川大で助手として働く．
同じ建物にいた轟木君と出会う．

自己紹介など

１５年前：横浜

神奈川大で助手として働く．
同じ建物にいた轟木君と出会う．

2007 **年〜**2015 **年：島根**

島根大学で働く．第二の修行時代．

自己紹介など

１５年前：横浜

神奈川大で助手として働く．
同じ建物にいた轟木君と出会う．

2007 年〜2015 年：島根

島根大学で働く．第二の修行時代．

３年前：インド

Ball **師匠「数学の学生の就職対策に再生核の理論はもってこい」**
カーネル法の数学的仕組みに詳しいことに気づく．

自己紹介など

１５年前：横浜

神奈川大で助手として働く．
同じ建物にいた轟木君と出会う．

2007 年〜2015 年：島根

島根大学で働く．第二の修行時代．

３年前：インド

Ball 師匠「数学の学生の就職対策に再生核の理論はもってこい」
カーネル法の数学的仕組みに詳しいことに気づく．

ここ数年：横須賀

機械学習について勉強中．しかし、応用は素人（通信空手黒帯のような状態）．

今回のお話

この話の内容

- 第１部　カーネル法とは何か？
- 第２部　カーネル法の理論と応用
- 第３部　サポートベクトルマシン入門

注意

学部２、３年生に講義するつもりで話します。

第一部　カーネル法とは何か？

Wikipedia によると

“カーネル関数$_1$ を使って、

計算複雑度の増大を抑えつつ内積にもとづく解析手法を

高次元特徴空間へ拡張$_2$ するアプローチを、

一般に カーネルトリック$_3$ と呼ぶ。”

まず，下線部 2 から解説します．

期末試験の問題

問題

次の積分の値を求めよ．

$$I = \int_{-\infty}^{\infty} e^{-x^2}\ dx$$

期末試験の問題

問題

次の積分の値を求めよ．

$$I = \int_{-\infty}^{\infty} e^{-x^2}\ dx$$

解答

$$\begin{aligned}
I^2 &= (\int_{-\infty}^{\infty} e^{-x^2}\ dx)(\int_{-\infty}^{\infty} e^{-y^2}\ dy) \\
&= \int_{\mathbb{R}^2} e^{-(x^2+y^2)}\ dxdy \\
&= \int_{\mathbb{R}^2} e^{-r^2} r\ drd\theta \\
&= \int_0^{2\pi}\ d\theta \int_0^{\infty} e^{-r^2} r\ dr = 2\pi[-e^{-r^2}/2]_0^{\infty} = \pi.
\end{aligned}$$

$$\therefore I = \sqrt{\pi}.$$

下線部 2 について

「高次元特徴空間へ拡張するアプローチを、・・・」

$\uparrow$

数学全体での基本的かつ重要なアイデア

- 高次元化することで，問題が簡単になることがある．
- 上空移行の原理（岡 潔）
- ジェットコースターの話（広中 平祐）
- $\displaystyle\int_{-\infty}^{\infty} \frac{1}{x^4+1}\, dx$ も $x \to z$ と変換することで求められる．

カーネル関数とは何か？

線型代数を思い出すと

$A = (a_{ij})$: $n \times n$ **対称行列（すなわち** $(a_{ji}) = (a_{ij})$**）**

$$A \geq 0 \overset{def}{\Leftrightarrow} \sum_{i,j=1}^{n} a_{ij} c_i c_j \geq 0 \quad (\forall (c_1, \ldots, c_n) \in \mathbb{R}^n).$$

- $A \geq 0 \Leftrightarrow \langle Ac, c \rangle_{\mathbb{R}^n} \geq 0 \ (\forall c \in \mathbb{R}^n) \Leftrightarrow A$ **の固有値が非負．**

カーネル関数とは何か？

設定

Ω: **集合（データはこの中の点）**

$K(x, y)$: $\Omega \times \Omega$ **上の関数**

カーネル関数とは何か？

設定

Ω: 集合（データはこの中の点）

$K(x, y)$: $\Omega \times \Omega$ 上の関数

カーネル関数

$K(x, y)$ が次の 1 と 2 をみたすとき，$K(x, y)$ をカーネル関数と呼ぶ．

1. $K(y, x) = K(x, y)$（対称性）
2. $\forall n \in \mathbb{N}$, $\forall w_1, \ldots, w_n \in \Omega$, $\forall c_1, \ldots, c_n \in \mathbb{R}$ に対し，

$$\sum_{i,j=1}^{n} K(w_i, w_j) c_i c_j \geq 0 \quad \text{（正定値性）}$$

例 1

$f(x)$: Ω 上の一変数関数

$$K(x,y) = f(x)f(y)$$

はカーネル関数．

なぜならば，

1. $K(y,x) = f(y)f(x) = f(x)f(y) = K(x,y)$
2. $$\sum_{i,j=1}^{n} K(w_i, w_j)c_i c_j = \sum_{i,j=1}^{n} f(w_i)f(w_j)c_i c_j = (\sum_{j=1}^{n} c_j f(w_j))^2 \geq 0.$$

例 2

$\Phi\colon \Omega \to \mathbb{R}^n$ **（写像）**

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathbb{R}^n}$$

はカーネル関数．

なぜならば，

1. $K(y, x) = \langle \Phi(y), \Phi(x) \rangle_{\mathbb{R}^n} = \langle \Phi(x), \Phi(y) \rangle_{\mathbb{R}^n} = K(x, y)$

2.

$$\begin{aligned} \sum_{i,j=1}^{n} K(w_i, w_j) c_i c_j &= \sum_{i,j=1}^{n} \langle \Phi(w_i), \Phi(w_j) \rangle_{\mathbb{R}^n} c_i c_j \\ &= \langle \sum_{i=1}^{n} c_i \Phi(w_i), \sum_{j=1}^{n} c_j \Phi(w_j) \rangle_{\mathbb{R}^n} \geq 0. \end{aligned}$$

統計学への応用

設定

$\{x_1, \ldots, x_n\} (\subset \Omega)$: **データの集合（有限個）**

現場からの要望

$\{x_1, \ldots, x_n\}$ **を適切な基準で二つに分割したい（例：健康診断）**

統計学への応用

設定

$\{x_1, \ldots, x_n\} (\subset \Omega)$: データの集合（有限個）

現場からの要望

$\{x_1, \ldots, x_n\}$ を適切な基準で二つに分割したい（例：健康診断）

数学的には

空間内に分布している点を（超）平面で分割できるか？

統計学への応用

設定

$\{x_1, \ldots, x_n\} (\subset \Omega)$: データの集合（有限個）

現場からの要望

$\{x_1, \ldots, x_n\}$ を適切な基準で二つに分割したい（例：健康診断）

数学的には

空間内に分布している点を（超）平面で分割できるか？

この問題の難しい点

データの分布が（超）平面と相性が良いとは限らない．

統計学への応用

設定

$\{x_1, \ldots, x_n\}(\subset \Omega)$: データの集合（有限個）

現場からの要望

$\{x_1, \ldots, x_n\}$ を適切な基準で二つに分割したい（例：健康診断）

数学的には

空間内に分布している点を（超）平面で分割できるか？

この問題の難しい点

データの分布が（超）平面と相性が良いとは限らない．

カーネルトリック

データ x_j を $k_{x_j} = K(\cdot, x_j)$ に変換せよ．$\Phi : x_j \mapsto k_{x_j}$（特徴写像）

カーネルトリックの数学的背景

K: Ω 上のカーネル関数 $(k_x := K(\cdot, x))$,

定理 (Moore-Aronszajn)

K には次をみたすヒルベルト空間 $\mathcal{H}_K$ がただ一つ対応する.

1. $\mathcal{H}_K$ は Ω 上の関数からなるヒルベルト空間.
2. $f(x) = \langle f, k_x \rangle_{\mathcal{H}_k}$ $(f \in \mathcal{H}_K, x \in \Omega)$.

カーネルトリックの数学的背景

K: Ω 上のカーネル関数 ($k_x := K(\cdot, x)$),

定理 (Moore-Aronszajn)

K には次をみたすヒルベルト空間 $\mathcal{H}_K$ がただ一つ対応する．

1. $\mathcal{H}_K$ は Ω 上の関数からなるヒルベルト空間．
2. $f(x) = \langle f, k_x \rangle_{\mathcal{H}_k}$ $(f \in \mathcal{H}_K, x \in \Omega)$.

用語の整理

- $\mathcal{H}_K$ は再生核ヒルベルト空間とよばれる．
- K はカーネル関数，$k_x = K(\cdot, x)$ は再生核．
- カーネル関数と再生核の関係：$K(x, y) = \langle k_y, k_x \rangle_{\mathcal{H}_K}$.

なぜ再生核ヒルベルト空間 (RKHS) を考えるのか

RKHS に期待される２つの機能

- **直交射影が使える．**
- **代入が内積で表される．**

なぜ再生核ヒルベルト空間 (RKHS) を考えるのか

RKHS に期待される２つの機能

- **直交射影が使える．**
- **代入が内積で表される．**

代入が内積で表される数学は良い数学

$$\int_{\Omega} f(x)g(x)\ dx \ \textbf{（連続）} \longleftrightarrow \sum_i a_i b_i \ \textbf{（離散）}$$

$$f(\lambda) = \frac{1}{2\pi i}\int_C \frac{f(z)}{z-\lambda}\ dz \quad \textbf{（コーシーの積分公式）},$$

$$f(a) = \int_{-\infty}^{\infty} f(x)\delta(x-a)\ dx \quad \textbf{（ディラックのデルタ関数）}.$$

第一部のまとめ

カーネル法（カーネルトリック）とは

- 非線型なデータを「直交射影」プラス「代入が内積（　積分）で表される仕組み」で扱う方法である．
- 特徴写像 $\Phi : x \mapsto k_x$ にデータの非線形性が組み込まれている（従って，問題は特徴写像の選び方（モデルの選択）である）．

$$\text{常微分方程式} \xrightarrow{\text{ラプラス変換}} \text{代数方程式}$$

$$\text{非線形なデータの問題} \xrightarrow{\text{カーネルトリック}} \text{線形代数の問題}$$

定理 (Aronszajn)

K_1, K_2 が Ω 上のカーネル関数ならば $K_1 + K_2$, $K_1 K_2$ も Ω 上のカーネル関数であり，$\mathcal{H}_{K_1+K_2}$, $\mathcal{H}_{K_1 K_2}$ を $\mathcal{H}_{K_1}$, $\mathcal{H}_{K_2}$ から構成できる．

定理 (Aronszajn)

K_1, K_2 が Ω 上のカーネル関数ならば $K_1 + K_2$, $K_1 K_2$ も Ω 上のカーネル関数であり，$\mathcal{H}_{K_1+K_2}$, $\mathcal{H}_{K_1 K_2}$ を $\mathcal{H}_{K_1}$, $\mathcal{H}_{K_2}$ から構成できる．

補足

カーネル関数の構成法（= RKHS の構成法）はたくさんある．

第二部　カーネル法の理論と応用

定理 (Aronszajn)

K_1, K_2 が Ω 上のカーネル関数ならば $K_1 + K_2$, $K_1 K_2$ も Ω 上のカーネル関数であり，$\mathcal{H}_{K_1+K_2}$, $\mathcal{H}_{K_1 K_2}$ を $\mathcal{H}_{K_1}$, $\mathcal{H}_{K_2}$ から構成できる．

補足

カーネル関数の構成法（= RKHS の構成法）はたくさんある．

カーネル法の言葉で言えば

新しい特徴写像（モデル）を次々に構成できる．
例：K がカーネル関数ならば e^K もカーネル関数（$\to$ ガウスカーネル）．

第二部　カーネル法の理論と応用

定理 (Aronszajn)

K_1, K_2 が Ω 上のカーネル関数ならば $K_1 + K_2$, $K_1 K_2$ も Ω 上のカーネル関数であり，$\mathcal{H}_{K_1+K_2}$, $\mathcal{H}_{K_1 K_2}$ を $\mathcal{H}_{K_1}$, $\mathcal{H}_{K_2}$ から構成できる．

補足

カーネル関数の構成法（= RKHS の構成法）はたくさんある．

カーネル法の言葉で言えば

新しい特徴写像（モデル）を次々に構成できる．
例：K がカーネル関数ならば e^K もカーネル関数（$\to$ ガウスカーネル）．

フォン ノイマン流の量子力学に詳しい方へ

RKHS は「ヒルベルト空間」と「自己共役作用素」の組

リプレゼンター定理１

問題１

$\{x_1, \ldots, x_n\} \subset \Omega$, $\{\lambda_1, \ldots, \lambda_n\} \subset \mathbb{R}$ **に対し，**

$$J(f) = \sum_{j=1}^{n} |f(x_j) - \lambda_j|^2$$

を最小化する $f \in \mathcal{H}_K$ **を見つけよ．**

リプレゼンター定理1

問題1

$\{x_1, \ldots, x_n\} \subset \Omega$, $\{\lambda_1, \ldots, \lambda_n\} \subset \mathbb{R}$ **に対し，**

$$J(f) = \sum_{j=1}^{n} |f(x_j) - \lambda_j|^2$$

を最小化する $f \in \mathcal{H}_K$ **を見つけよ．**

準備

P **を** $\{k_{x_1}, \ldots k_{x_n}\}$ **で張られる空間への直交射影とすると，**

$$Pf = \sum_{j=1}^{n} c_j k_{x_j}$$

と表される（ヒルベルト空間を考える御利益）．

リプレゼンター定理１

解法

$$f(x_i) = \langle f, k_{x_i} \rangle_{\mathcal{H}_K} = \langle Pf, k_{x_i} \rangle_{\mathcal{H}_K} = \langle \sum_{j=1}^{n} c_j k_{x_j}, k_{x_i} \rangle_{\mathcal{H}_K}$$

から

$$\begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix} = (K(x_i, x_j)) \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

が導かれる．この右辺を Kc と表せば，

$$J(f) = \sum_{j=1}^{n} |f(x_j) - \lambda_j|^2 = \|Kc - \lambda\|_{\mathbb{R}^n}^2.$$

（関数 f の問題がベクトル c の問題になった）

カーネル法の簡単な例

問題２

$\{x_1, \ldots, x_n\} \subset \mathbb{R}$, $\{\lambda_1, \ldots, \lambda_n\} \subset \mathbb{R}$ **に対し，**

$$J(f) = \sum_{j=1}^{n} |p(x_j) - \lambda_j|^2$$

を最小化する多項式 p $(d = \deg p < n - 1)$ **を見つけよ．**

カーネル法の簡単な例

問題２

$\{x_1, \ldots, x_n\} \subset \mathbb{R}$, $\{\lambda_1, \ldots, \lambda_n\} \subset \mathbb{R}$ **に対し，**

$$J(f) = \sum_{j=1}^{n} |p(x_j) - \lambda_j|^2$$

を最小化する多項式 p $(d = \deg p < n-1)$ **を見つけよ．**

方針

カーネル法により問題１に帰着させる
（次数 d **以下の多項式全体からなる** $\mathcal{H}_K$ **を見つけてくればよい）．**

カーネル法の簡単な例

解法

特徴写像として

$$\Phi : \{x_1, \ldots, x_n\} \to \mathcal{P} := (d\ \textbf{次以下の多項式全体}) \cong \mathbb{R}^{d+1}$$
$$x_j \mapsto 1 + x_j x + \cdots + x_j^d x^d$$

を採用する．

$\mathcal{P}$ **には次のようにして再生核ヒルベルト空間の構造が入る．**

$p(x) = a_0 + a_1 x + \cdots + a_d x^d$ **に対し，**

$$\langle p, \Phi(x_j)\rangle_{\mathcal{P}} := \left\langle \begin{pmatrix} a_0 \\ \vdots \\ a_d \end{pmatrix}, \begin{pmatrix} 1 \\ \vdots \\ x_j^d \end{pmatrix} \right\rangle_{\mathbb{R}^{d+1}}$$
$$= a_0 + a_1 x_j + \cdots + a_d x_j^d.$$

リプレゼンター定理２

- P を $\{k_{x_1}, \ldots k_{x_n}\}$ で張られる空間への直交射影とすれば，

$$(Pf)(x_i) = f(x_i) \quad (i = 1, \ldots, n).$$

- $f(x_i) = 0 \ (i = 1, \ldots, n)$ となる f は P で除去できる．
- $\|f\|^2_{\mathcal{H}_K} = \|Pf\|^2_{\mathcal{H}_K} + \|(I - P)f\|^2_{\mathcal{H}_K}$ （三平方の定理）．
- $f(x_i)$ に関する問題は $(Pf)(x_i)$ を考えればよい（問題を有限次元に落とせる）．

第二部のまとめ

カーネル法勉強の目安

- 内積の計算ができて有名な定理の意味がわかれば基本は OK．
- カーネル関数のいろいろな構成法を知っておくと将来便利かも．

参考文献

[1] 赤穂昭太郎，カーネル多変量解析，岩波書店．

[2] 竹内一郎，烏山昌幸，サポートベクトルマシン，講談社．

[3] 金森敬文，統計的学習理論，講談社．

[4] 福水健次，カーネル法入門，朝倉書店．

[5] C. M. ビショップ，パターン認識と機械学習，丸善出版．

[6] 私の講義ノート，https://researchmap.jp/mseto/ の資料公開．

第三部　サポートベクトルマシン入門編

設定

$D = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^2$ **（データの集合）**

各データには符号 $\lambda_j \in \{-1, +1\}$ **がラベル付けされている．**

すなわち， D **は**

$$D_+ = \{(x_j, y_j) \in D : \lambda_j = +1\}, \quad D_- = \{(x_j, y_j) \in D : \lambda_j = -1\}$$

と分割される．

第三部　サポートベクトルマシン入門編

設定

$D = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^2$　（データの集合）

各データには符号 $\lambda_j \in \{-1, +1\}$ がラベル付けされている．

すなわち，D は

$$D_+ = \{(x_j, y_j) \in D : \lambda_j = +1\}, \quad D_- = \{(x_j, y_j) \in D : \lambda_j = -1\}$$

と分割される．

問題

$$D_+ \subset \{(x, y) \in \mathbb{R}^2 : f(x, y) > 0\}, \quad D_- \subset \{(x, y) \in \mathbb{R}^2 : f(x, y) < 0\}$$

をみたす適切な関数 $f(x, y)$ を見つけよ．

サポートベクトルマシン入門編

問題

$$D_+ \subset \{(x, y) \in \mathbb{R}^2 : f(x, y) > 0\}, \quad D_- \subset \{(x, y) \in \mathbb{R}^2 : f(x, y) < 0\}$$

をみたす適切な関数 $f(x, y)$ を見つけよ．

コメント

- **適切な関数とは簡単なものであってほしい（過学習の問題）．**
- **D_+ と D_- の間に原点を通る直線が引けるとき，D は線形分離できるという（最も単純な場合）．**
- **データの分布が線形分離と相性が悪いときどうする？**

拡張された問題

$$D_+ \subset \{(x, y) \in \mathbb{R}^2 : f(x, y) > 0\}, \quad D_- \subset \{(x, y) \in \mathbb{R}^2 : f(x, y) < 0\}$$

をみたす関数 $f(x, y) = a + bx + cy + dx^2 + ey^2$ **を見つけよ．**

サポートベクトルマシン入門編

拡張された問題

$$D_+ \subset \{(x,y) \in \mathbb{R}^2 : f(x,y) > 0\}, \quad D_- \subset \{(x,y) \in \mathbb{R}^2 : f(x,y) < 0\}$$

をみたす関数 $f(x,y) = a + bx + cy + dx^2 + ey^2$ **を見つけよ．**

Step 1

写像

$$\Phi : \mathbb{R}^2 \to \mathbb{R}^5 \quad \Phi(x,y) = (1, x, y, x^2, y^2)^t$$

を特徴写像として採用し，$\mathbb{R}^2$ **上のカーネル関数を**

$$k((x,y),(z,w)) = \langle \Phi(x,y), \Phi(z,w) \rangle_{\mathbb{R}^5} \quad ((x,y),(z,w) \in \mathbb{R}^2)$$

と定める．

サポートベクトルマシン入門編

Step 2

このとき，$\boldsymbol{v} = (a, b, c, d, e)^t \in \mathbb{R}^5$ に対し，

$$\langle \boldsymbol{v}, \Phi(x_j, y_j) \rangle_{\mathbb{R}^5} = a + bx_j + cy_j + dx_j^2 + ey_j^2.$$

よって，$\mathbb{R}^5$ のベクトル $\boldsymbol{v}$ で

$$\Phi(D_+) \subset \{\boldsymbol{x} \in \mathbb{R}^5 : \langle \boldsymbol{v}, \boldsymbol{x} \rangle_{\mathbb{R}^5} > 0\}, \quad \Phi(D_-) \subset \{\boldsymbol{x} \in \mathbb{R}^5 : \langle \boldsymbol{v}, \boldsymbol{x} \rangle_{\mathbb{R}^5} < 0\}$$

をみたすものを見つけてくればよい．

Step 2

このとき，$\boldsymbol{v} = (a, b, c, d, e)^t \in \mathbb{R}^5$ に対し，

$$\langle \boldsymbol{v}, \Phi(x_j, y_j) \rangle_{\mathbb{R}^5} = a + bx_j + cy_j + dx_j^2 + ey_j^2.$$

よって，$\mathbb{R}^5$ のベクトル $\boldsymbol{v}$ で

$$\Phi(D_+) \subset \{\boldsymbol{x} \in \mathbb{R}^5 : \langle \boldsymbol{v}, \boldsymbol{x} \rangle_{\mathbb{R}^5} > 0\}, \quad \Phi(D_-) \subset \{\boldsymbol{x} \in \mathbb{R}^5 : \langle \boldsymbol{v}, \boldsymbol{x} \rangle_{\mathbb{R}^5} < 0\}$$

をみたすものを見つけてくればよい．

Step 3

$$V = \{\boldsymbol{x} \in \mathbb{R}^5 : \langle \boldsymbol{v}, \boldsymbol{x} \rangle_{\mathbb{R}^5} = 0\}$$

は $\mathbb{R}^5$ 内の原点を通る超平面である（$\mathbb{R}^5$ での線形分離の問題に帰着された）．また，リプレゼンター定理により，$\boldsymbol{v} = \sum_{j=1}^{n} c_j \Phi(x_j, y_j)$ と仮定してよいことに注意しよう．

例題（簡単ですが）

例題 1

$\mathbb{R}^3$ 内の D_+ と D_- が平面 $ax+by+cz+d=0$ で分離されるとき，$\Phi(D_+)$ と $\Phi(D_-)$ が $\mathbb{R}^4$ で線形分離できるような特徴写像 Φ を定めよ．

例題（簡単ですが）

例題 1

$\mathbb{R}^3$ 内の D_+ と D_- が平面 $ax+by+cz+d=0$ で分離されるとき，$\Phi(D_+)$ と $\Phi(D_-)$ が $\mathbb{R}^4$ で線形分離できるような特徴写像 Φ を定めよ．

略解

$$D_+ \subset \{(x,y,z)^t \in \mathbb{R}^3 : ax+by+cz+d>0\}$$

と仮定する．D_- についても同様．特徴写像として，

$$\Phi : \mathbb{R}^3 \to \mathbb{R}^4, \quad (x,y,z)^t \mapsto (x,y,z,1)^t$$

を採用すると，$(x,y,z)^t \in D_+$ と $v=(a,b,c,d)^t$ に対し，

$$\langle v, \Phi(x,y,z)\rangle_{\mathbb{R}^4} = ax+by+cz+d>0$$

が成り立つ．よって，

$$\Phi(D_+) \subset \{(x,y,z,w)^t \in \mathbb{R}^4 : \langle v, (x,y,z,w)^t\rangle_{\mathbb{R}^4} > 0\}.$$

例題（簡単ですが）

例題２

$\mathbb{R}^2$ 内の D_+ と D_- が円 $x^2 + y^2 = 1$ で分離されるとき，$\Phi(D_+)$ と $\Phi(D_-)$ が $\mathbb{R}^3$ で線形分離できるような特徴写像 Φ を定めよ．

例題（簡単ですが）

例題２

$\mathbb{R}^2$ 内の D_+ と D_- が円 $x^2 + y^2 = 1$ で分離されるとき，$\Phi(D_+)$ と $\Phi(D_-)$ が $\mathbb{R}^3$ で線形分離できるような特徴写像 Φ を定めよ．

略解

$$D_+ \subset \{(x, y)^t \in \mathbb{R}^2 : x^2 + y^2 > 1\}$$

と仮定しよう．D_- についても同様．特徴写像として，

$$\Phi : \mathbb{R}^2 \to \mathbb{R}^3, \quad (x, y)^t \mapsto (x, y, x^2 + y^2 - 1)^t$$

を採用すると，$(x, y)^t \in D_+$ と $v = (0, 0, 1)^t$ に対し，

$$\langle v, \Phi(x, y)\rangle_{\mathbb{R}^4} = x^2 + y^2 - 1 > 0$$

が成り立つ．よって，

$$\Phi(D_+) \subset \{(x, y, z)^t \in \mathbb{R}^3 : \langle v, (x, y, z)^t\rangle_{\mathbb{R}^3} > 0\}.$$

サポートベクトルマシン入門編

ハードマージン法

$D = D_+ \cup D_-$ が線形分離できるとき，

$$\max_{H} \min_{1 \le i \le n} d(x_i, H) \quad (D = \{x_1, \ldots, x_n\})$$

を解いて，D_+ と D_- のちょうど中間にある平面 H を選ぶ方法．

ソフトマージン法

線形分離できない場合の次善策．

- 角度と距離の問題は内積の問題である（カーネル法を使った先でも角度と距離が考えられる）．
- カーネル法を使っても，リプレゼンター定理により，n 変数の２次関数の問題になる（n はデータ数）．