

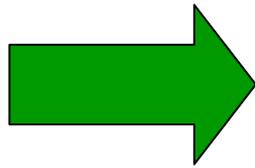
# 準教師付き局所フィッシャー 判別分析による次元削減



杉山	将	(東京工業大学)
井手	剛	(IBM)
中島	伸一	(NIKON)
瀬々	潤	(お茶の水女子大学)

# 次元削減

- **次元の呪い**: 高次元データは扱いにくい



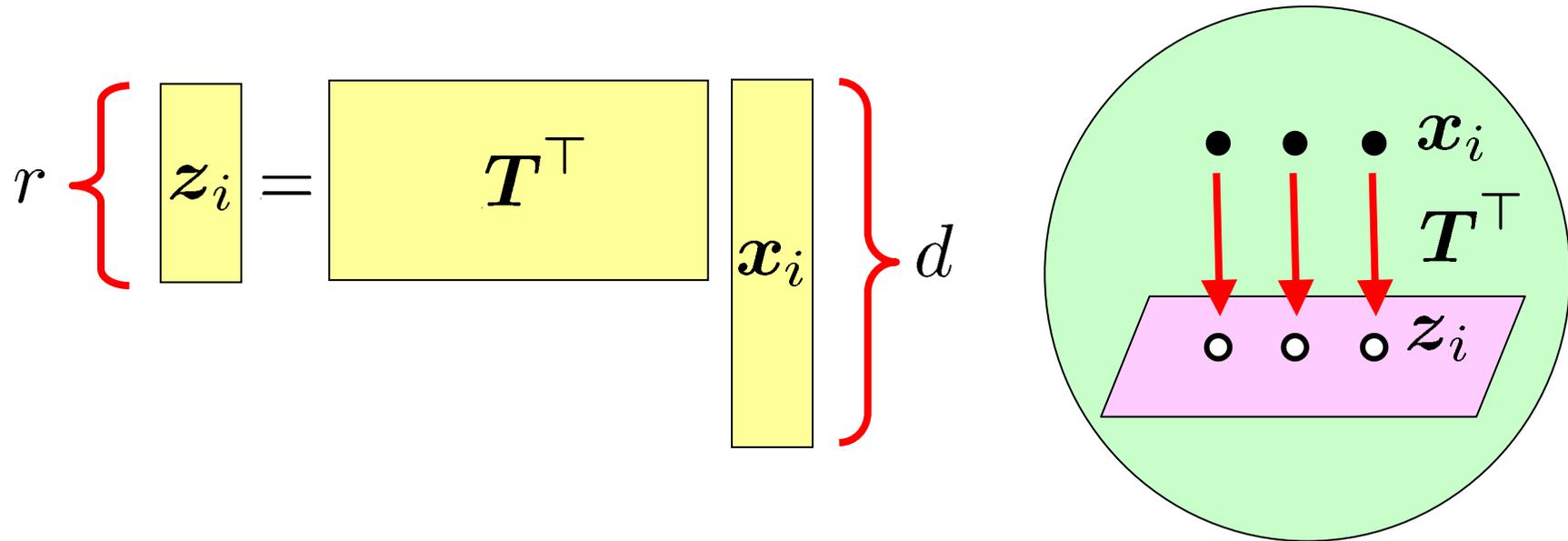
データの本質的な構造を保ちつつ  
次元数を削減したい



# 線形次元削減

## ■ 本発表では線形の次元削減問題を考える

- 高次元標本  $\{x_i\}_{i=1}^n$   $x_i \in \mathbb{R}^d$
- 埋め込み行列  $T$
- 埋め込まれた低次元標本  $\{z_i\}_{i=1}^n$   $z_i \in \mathbb{R}^r$



## ■ 適切な埋め込み行列 $T$ を求めたい!

# 本発表の構成

4

1. 線形次元削減問題
2. 主成分分析(PCA)
3. 局所性保存射影(LPP)
4. フィッシャー判別分析(FDA)
5. 局所フィッシャー判別分析(LFDA)
6. 準教師つき局所フィッシャー判別分析(SELF)
7. 次元削減と計量学習

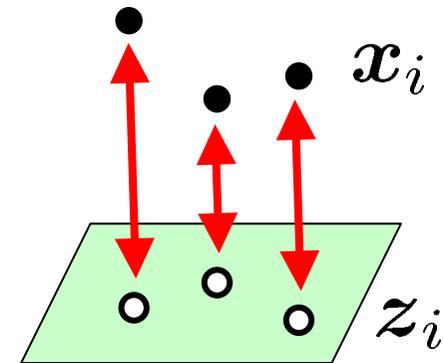


# 主成分分析(PCA)

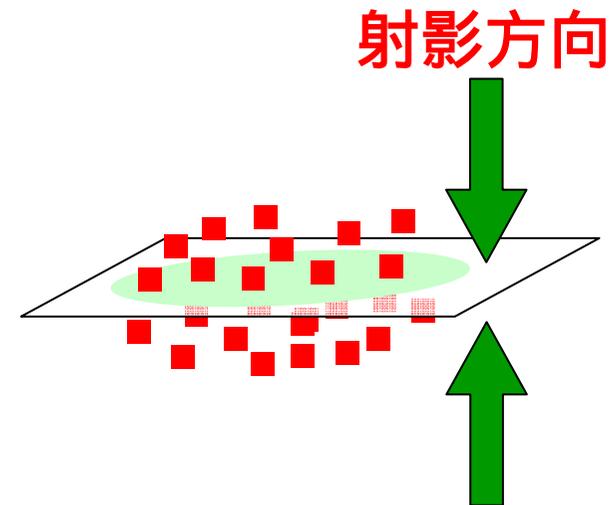
5

## ■ 基本アイデア:

- 標本の**最良近似**を与える  
**低次元部分空間**を求める



- 埋め込み後の標本の散らばりが最大になる部分空間を求めればよい!



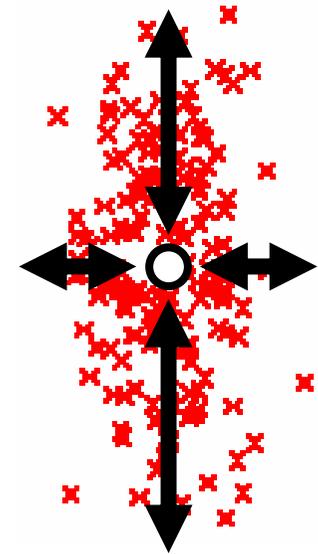
# PCA (続き)

6

## ■ 全標本散布行列:

$$S^{(t)} = \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top$$

$$\boldsymbol{\mu} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$



## ■ PCA規準: 埋め込み後の全標本散布を大きく

$$\max_T \left[ \text{tr}(\mathbf{T}^\top \mathbf{S}^{(t)} \mathbf{T} \underbrace{(\mathbf{T}^\top \mathbf{T})^{-1}}_{\text{正規化}}) \right]$$

## ■ 解: 全標本散布行列 $S^{(t)}$ の上位の固有ベクトルを並べる

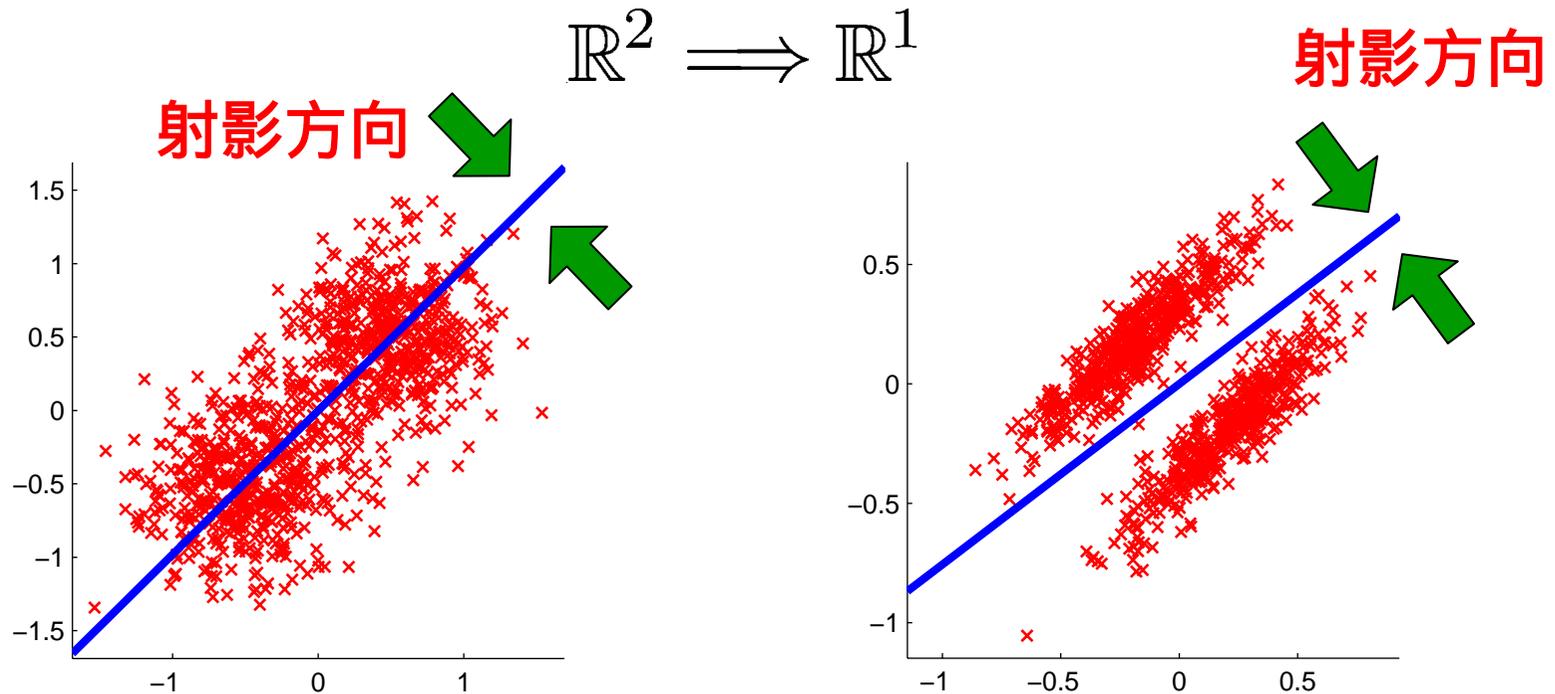
$$\mathbf{T}_{PCA} = (\boldsymbol{\varphi}_1 | \boldsymbol{\varphi}_2 | \cdots | \boldsymbol{\varphi}_r)$$

$$\mathbf{S}^{(t)} \boldsymbol{\varphi} = \lambda \boldsymbol{\varphi}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$$

# PCAの例

7



- データの**大域的な構造**が保存される。
- クラスタ等の**局所的な構造**は保存されるとは限らない。

# 本発表の構成

8

1. 線形次元削減問題
2. 主成分分析(PCA)
3. 局所性保存射影(LPP)
4. フィッシャー判別分析(FDA)
5. 局所フィッシャー判別分析(LFDA)
6. 準教師つき局所フィッシャー判別分析(SELF)
7. 次元削減と計量学習

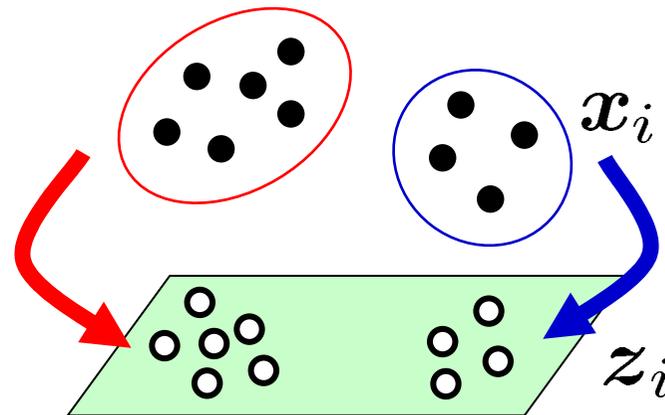


# 局所性保存射影(LPP)

9

He & Niyogi (NIPS2003)

- **基本アイデア**: “似ている” 標本同士を**近くに**埋め込む



➡ データの**局所的な構造**が保存される

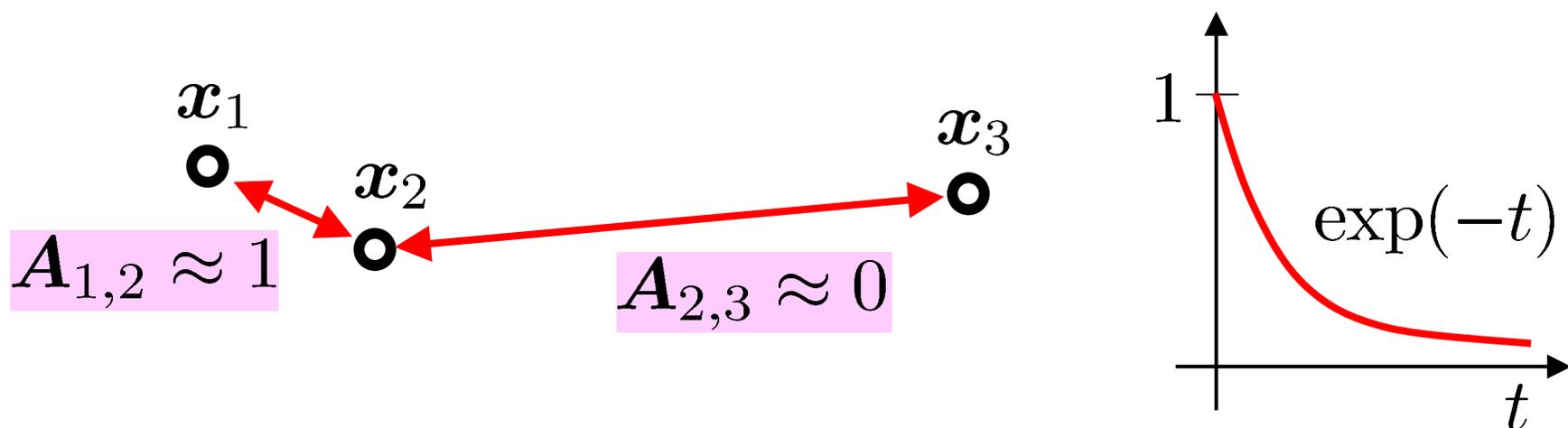
# LPP (続き)

10

## ■ 類似度行列:

- 近くにある標本同士の類似度は大きい
- 遠くにある標本同士の類似度は小さい
- 例:

$$A_{i,j} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right)$$



# LPP (続き)

11

## ■ 局所性行列:

$A_{i,j}$ : 類似度

$$S^{(l)} = \frac{1}{2n} \sum_{i,j=1}^n A_{i,j} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top$$

## ■ LPP規準: 類似度 $A_{i,j}$ が大きい標本同士を近づける

$$\min_T \left[ \text{tr}(\mathbf{T}^\top \mathbf{S}^{(l)} \mathbf{T} \underbrace{(\mathbf{T}^\top \mathbf{T})^{-1}}_{\text{正規化}}) \right]$$

## ■ 解: 局所性行列 $S^{(l)}$ の下位の固有ベクトルを並べる

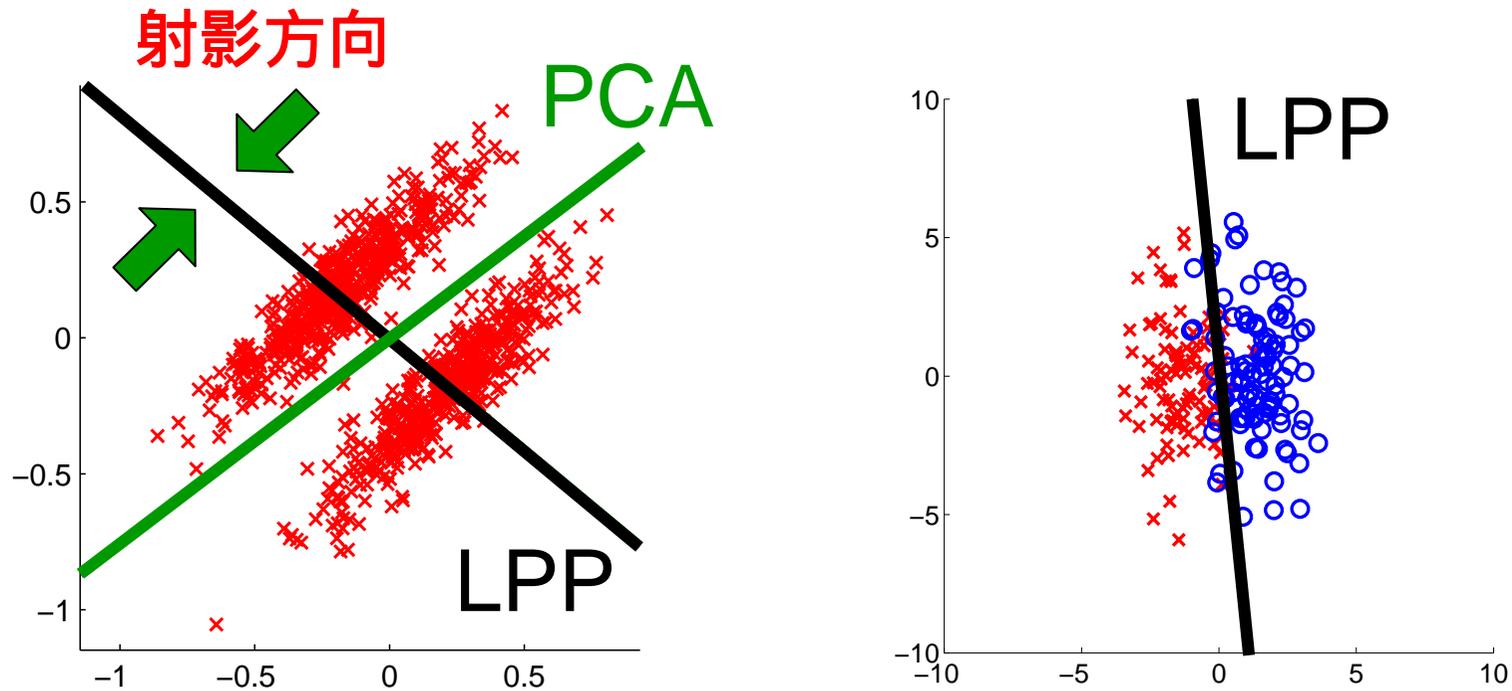
$$\mathbf{T}_{LPP} = (\varphi_d | \varphi_{d-1} | \cdots | \varphi_{d-r+1})$$

$$S^{(l)} \varphi = \lambda \varphi$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$$

# LPPの例

12



- クラスタ構造が保存される。
- 教師なしのため、クラス間の分離性は考慮されない。

# 本発表の構成

1. 線形次元削減問題
2. 主成分分析(PCA)
3. 局所性保存射影(LPP)
4. **フィッシャー判別分析(FDA)**
5. 局所フィッシャー判別分析(LFDA)
6. 準教師つき局所フィッシャー判別分析(SELF)
7. 次元削減と計量学習



# 教師付き次元削減

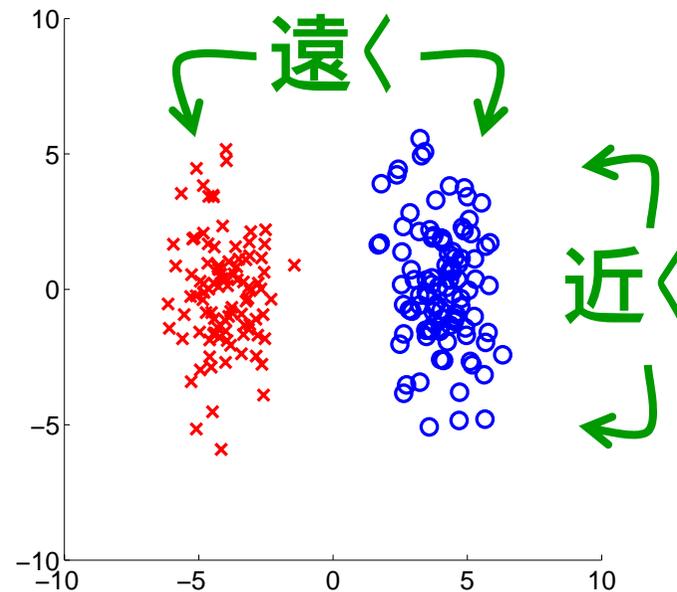
## ■ 教師つき学習:

- ラベルあり標本

$$\{(x_i, y_i)\}_{i=1}^n$$

$$y_i \in \{1, 2, \dots, c\}$$

- 同じクラスの標本は**近く**にまとめたい
- 違うクラスの標本は**遠く**に離したい



# フィッシャー判別分析(FDA)

15

## ■ クラス内散布行列:

$$S^{(w)} = \sum_{m=1}^c \sum_{i:y_i=m} (x_i - \mu_m)(x_i - \mu_m)^T$$

$$\mu_m = \frac{1}{n_m} \sum_{i:y_i=m} x_i$$

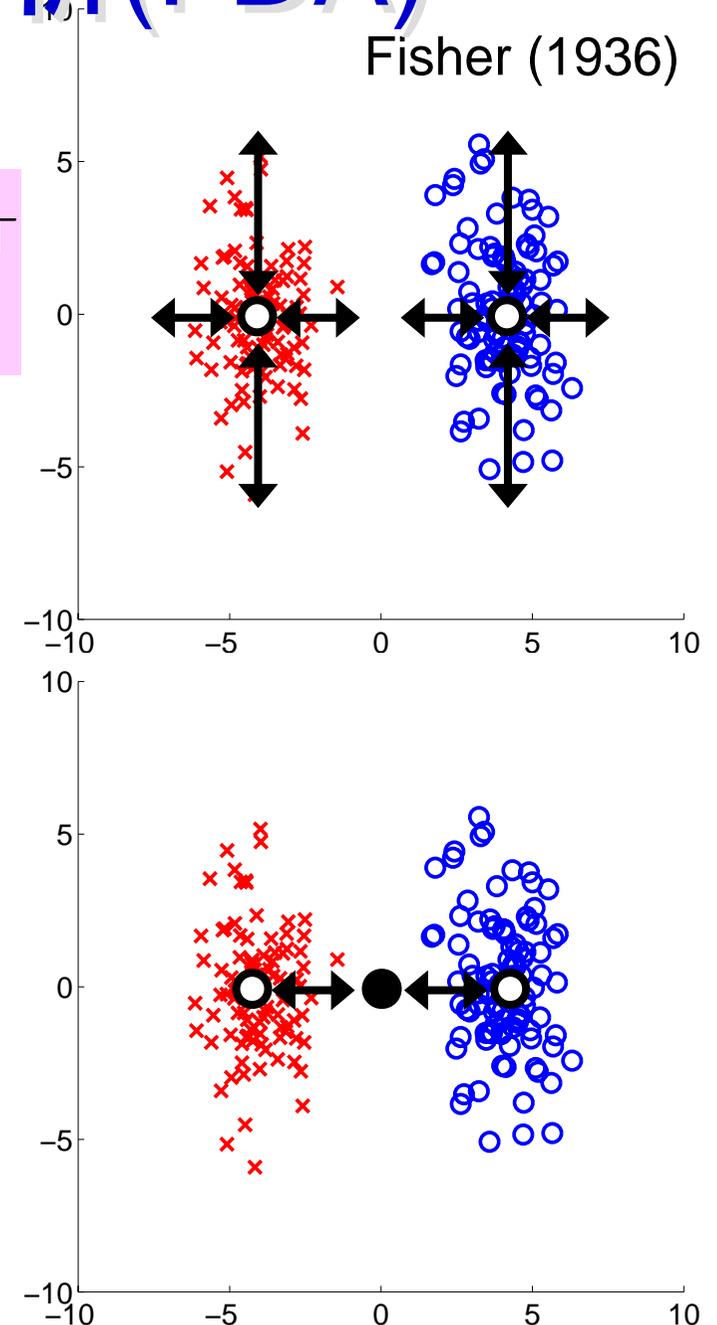
$n_m$ : クラス  $m$  の標本数

## ■ クラス間散布行列:

$$S^{(b)} = \sum_{m=1}^c n_m (\mu_m - \mu)(\mu_m - \mu)^T$$

$$\mu = \frac{1}{n} \sum_i x_i$$

$n$ : 全標本数



# FDA (続き)

16

## ■ FDA規準:

- クラス間散布を大きく
- クラス内散布を小さく

$$\max_T \left[ \text{tr}(\mathbf{T}^\top \mathbf{S}^{(b)} \mathbf{T} (\mathbf{T}^\top \mathbf{S}^{(w)} \mathbf{T})^{-1}) \right]$$

- ## ■ 解: クラス内・クラス間散布行列の一般化 固有値問題の上位の固有ベクトルを並べる

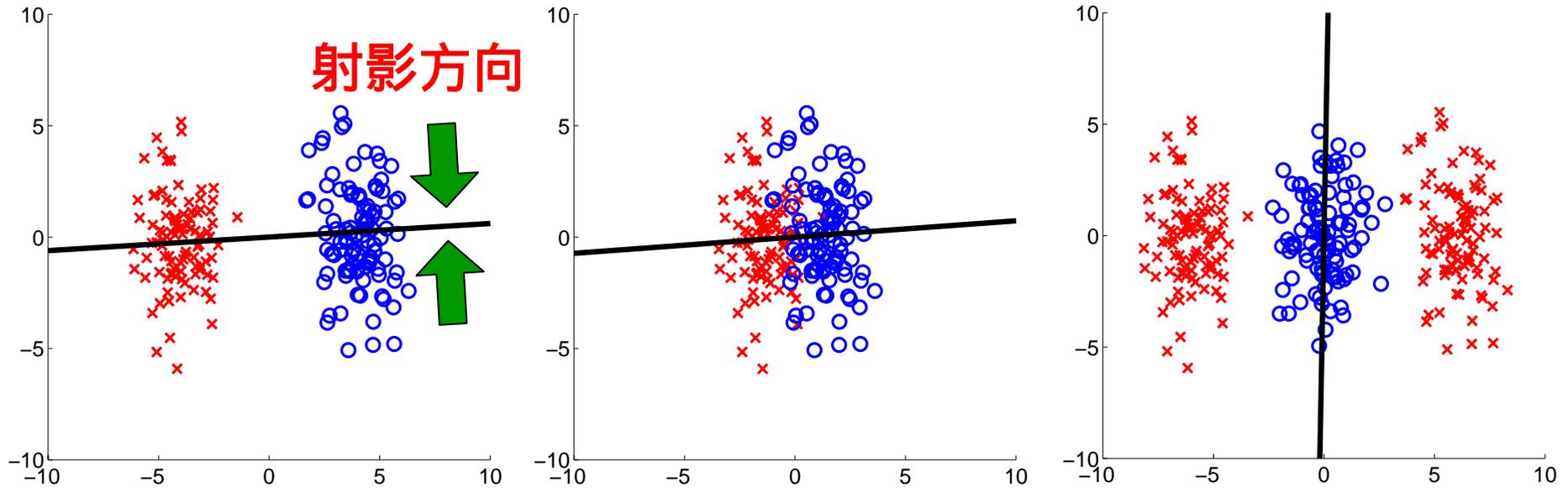
$$\mathbf{T}_{FDA} = (\varphi_1 | \varphi_2 | \cdots | \varphi_r)$$

$$\mathbf{S}^{(b)} \varphi = \lambda \mathbf{S}^{(w)} \varphi$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$$

# FDAの例

17



- 異なるクラスの標本がうまく分離される
- クラス内に多峰構造がある場合はうまくいかない
- $\text{rank}(S^{(b)}) = c - 1$  のため、最大でも  $c - 1$  個の特徴しか取り出せない

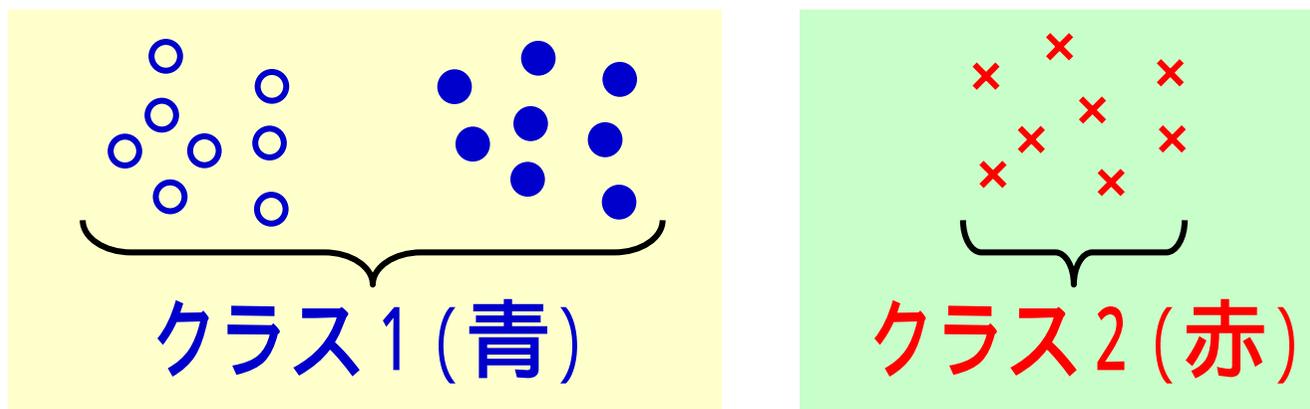
$c$  : クラス数

# 本発表の構成

1. 線形次元削減問題
2. 主成分分析(PCA)
3. 局所性保存射影(LPP)
4. フィッシャー判別分析(FDA)
5. 局所フィッシャー判別分析(LFDA)
6. 準教師つき局所フィッシャー判別分析(SELF)
7. 次元削減と計量学習



# クラス内多峰性



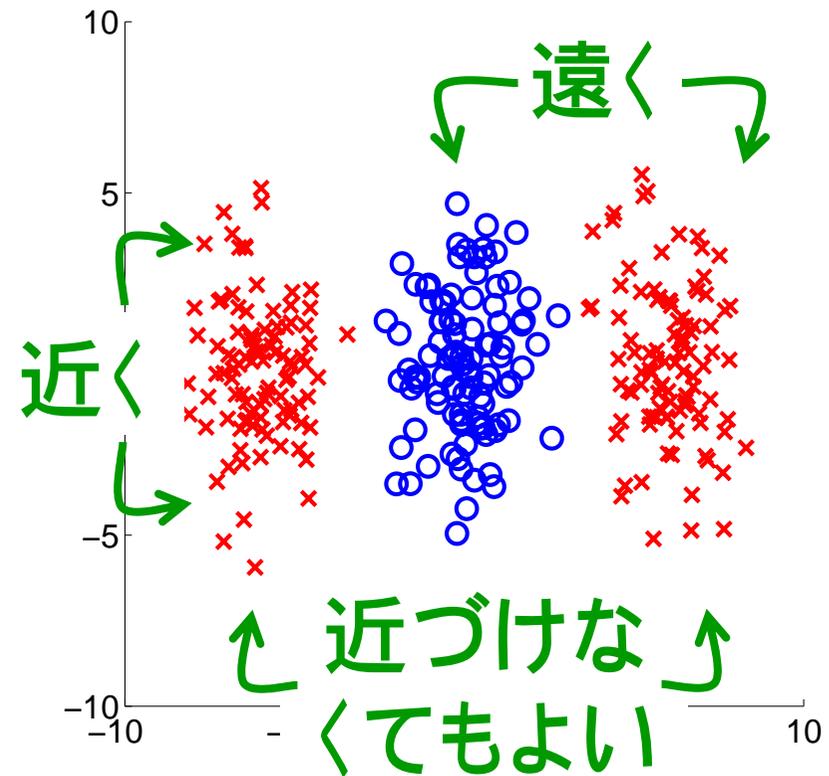
- あるクラスのデータが**多峰性**を持つ
  - **医療データ:**  
ホルモンアンバランス(多い・少ない) vs. 正常
  - **数字認識:**  
偶数(0, 2, 4, 6, 8) vs. 奇数(1, 3, 5, 7, 9)
  - **多クラス分類:**  
一クラス vs. 残りのクラス (one-versus-rest)

# 局所フィッシャー判別分析(LFDA) <sup>20</sup>

Sugiyama (JMLR2007)

## ■ 基本アイデア:

- 同じクラスの**近く**の標本は**近くに埋め込む**
- 同じクラスでも**遠く**の標本は**近づけなくてもよい**
- 異なるクラスの標本は**遠く**に埋め込む



LPPとFDAを組み合わせる！

# FDAの散布行列のペアワイズ表現<sup>21</sup>

$$\max_T \left[ \text{tr}(\mathbf{T}^\top \mathbf{S}^{(b)} \mathbf{T} (\mathbf{T}^\top \mathbf{S}^{(w)} \mathbf{T})^{-1}) \right]$$

$$\mathbf{S}^{(w)} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{i,j}^{(w)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top$$

$$\mathbf{W}_{i,j}^{(w)} = \begin{cases} 1/n_{y_i} & (y_i = y_j) \\ 0 & (y_i \neq y_j) \end{cases}$$

$$\mathbf{S}^{(b)} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{i,j}^{(b)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top$$

$$\mathbf{W}_{i,j}^{(b)} = \begin{cases} 1/n - 1/n_{y_i} & (y_i = y_j) \\ 1/n & (y_i \neq y_j) \end{cases}$$

同じクラスの標本は近くに埋め込む

違うクラスの標本は遠くに埋め込む

# LFDA (続き)

## ■ 局所クラス内散布行列:

$A_{i,j}$ : 類似度

$$S^{(lw)} = \frac{1}{2} \sum_{i,j=1}^n W_{i,j}^{(lw)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top$$

$$W_{i,j}^{(lw)} = \begin{cases} A_{i,j}/n_{y_i} & (y_i = y_j) \\ 0 & (y_i \neq y_j) \end{cases}$$

## ■ 局所クラス間散布行列:

$$S^{(lb)} = \frac{1}{2} \sum_{i,j=1}^n W_{i,j}^{(lb)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top$$

$$W_{i,j}^{(lb)} = \begin{cases} A_{i,j}(1/n - 1/n_{y_i}) & (y_i = y_j) \\ 1/n & (y_i \neq y_j) \end{cases}$$

## ■ $A_{i,j} = 1$ のとき, 元のクラス内・クラス間散布と一致

## ■ LFDA規準:

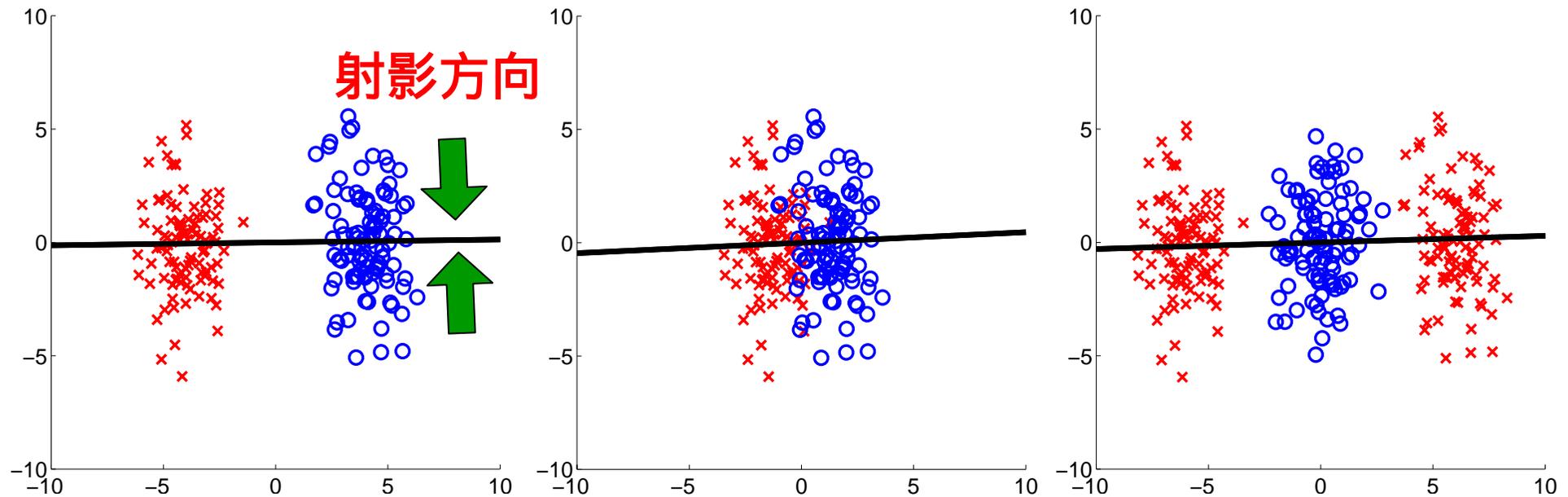
- 局所クラス間散布を大きく
- 局所クラス内散布を小さく

$$\max_T \left[ \text{tr}(\mathbf{T}^\top \mathbf{S}^{(lb)} \mathbf{T} (\mathbf{T}^\top \mathbf{S}^{(lw)} \mathbf{T})^{-1}) \right]$$

- ## ■ 解: 局所クラス内・クラス間散布行列の一般化固有値問題の上位の固有ベクトルを並べる

$$\mathbf{S}^{(lb)} \varphi = \lambda \mathbf{S}^{(lw)} \varphi$$

# LFDAの例



- クラス間の分離性が保存される。
- クラス内のクラスタ構造も保存される。
- 通常は  $\text{rank}(S^{(lb)}) \gg c$  が成り立つため、  
任意の個数の特徴が取り出せる

$c$  : クラス数

# 甲状腺疾患データの可視化

## ■ 甲状腺疾患データ (5次元)

- T3-resin uptake test.
- Total Serum thyroxin as measured by the isotopic displacement method.

など

## ■ ラベル: **正常** or **異常**

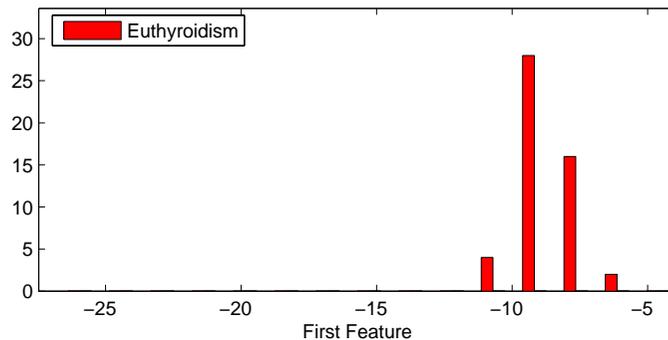
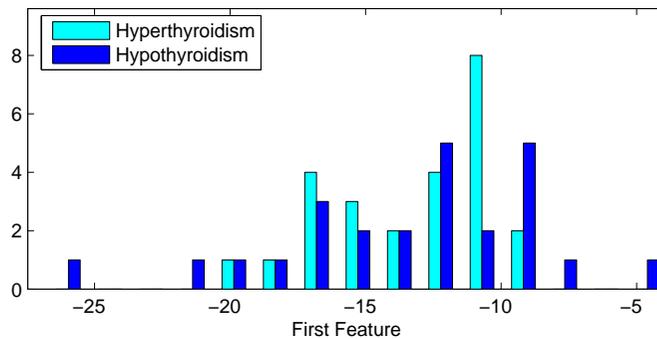
## ■ 甲状腺異常には

- **機能亢進 (こうしん)**: 機能が強すぎる
- **機能低下**: 機能が弱すぎる

の2種類がある

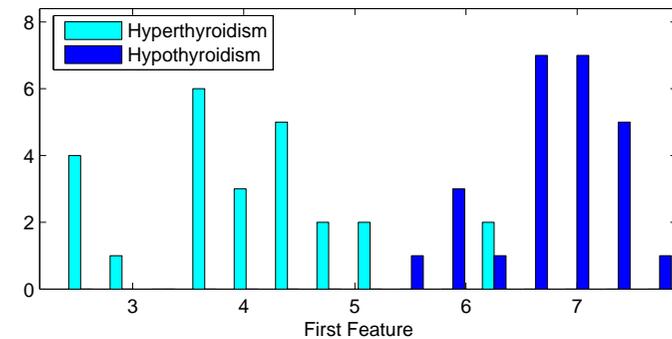
# 可視化結果(一次元)

## FDA

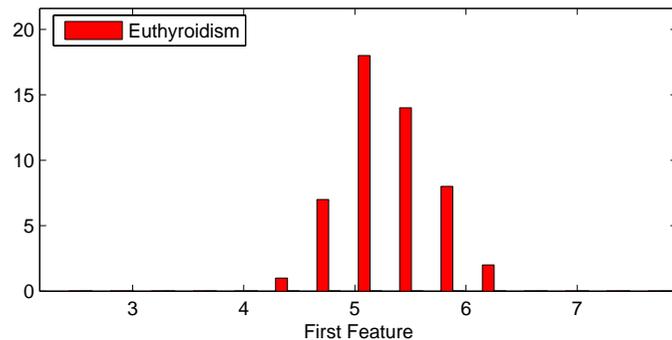


- 正常と異常はうまく分かれる
- 機能亢進と低下は混ざってしまう

## LFDA



異常



正常

- 正常と異常はうまく分かれる
- 機能亢進と低下もうまく分かれる
- 見つかった特徴は甲状腺機能のレベルと強い負の相関

# 次元削減 + 最近傍識別の誤認識率<sup>27</sup>

	LFDA	LDI	NCA	MCML	LPP	PCA
banana	13.7(0.8)	13.6(0.8)	14.3(2.0)	39.4(6.7)	13.6(0.8)	13.6(0.8)
b-cancer	34.7(4.3)	36.4(4.9)	34.9(5.0)	34.0(5.8)	33.5(5.4)	34.5(5.0)
diabetes	32.0(2.5)	30.8(1.9)		31.2(2.1)	31.5(2.5)	31.2(3.0)
f-solar	39.2(5.0)	39.3(4.8)			39.2(4.9)	39.1(5.1)
german	29.9(2.8)	30.7(2.4)	29.8(2.6)	31.3(2.4)	30.7(2.4)	30.2(2.4)
heart	21.9(3.7)	23.9(3.1)	23.0(4.3)	23.3(3.8)	23.3(3.8)	24.3(3.5)
image	3.2(0.8)	3.0(0.6)		4.7(0.8)	3.6(0.7)	3.4(0.5)
ringnorm	21.1(1.3)	17.5(1.0)	21.8(1.3)	22.0(1.2)	20.6(1.1)	21.6(1.4)
splice	16.9(0.9)	17.9(0.8)		17.3(0.9)	23.2(1.2)	22.6(1.3)
thyroid	4.6(2.6)	8.0(2.9)	4.5(2.2)	18.5(3.8)	4.2(2.9)	4.9(2.6)
titanic	33.1(11.9)	33.1(11.9)	33.0(11.9)	33.1(11.9)	33.0(11.9)	33.0(12.0)
twonorm	3.5(0.4)	4.1(0.6)	3.7(0.6)	3.5(0.4)	3.7(0.7)	3.6(0.6)
waveform	12.5(1.0)	20.7(2.5)	12.6(0.8)	17.9(1.5)	12.4(1.0)	12.7(1.2)
計算時間比	1.00	1.11	97.23	70.61	1.04	0.91

- 次元数はクロスバリデーションで決定, 数値は平均(標準偏差)
- 青地はクラス内多峰性があるデータ, 赤字は5%のt検定で優位に良いもの
- LDI: 局所識別情報 (Hastie & Tibshirani, IEEE-PAMI1996)
- NCA: 近傍成分分析 (Goldberger, Roweis, Hinton & Salakhutdinov, NIPS2004)
- MCML: 最大縮退計量学習 (Globerson & Roweis, NIPS2005)

# 本発表の構成

1. 線形次元削減問題
2. 主成分分析(PCA)
3. 局所性保存射影(LPP)
4. フィッシャー判別分析(FDA)
5. 局所フィッシャー判別分析(LFDA)
6. 準教師つき局所フィッシャー判別分析(SELF)
7. 次元削減と計量学習



# 準教師つき次元削減

## ■ 準教師つき学習:

- 少数のラベルあり標本

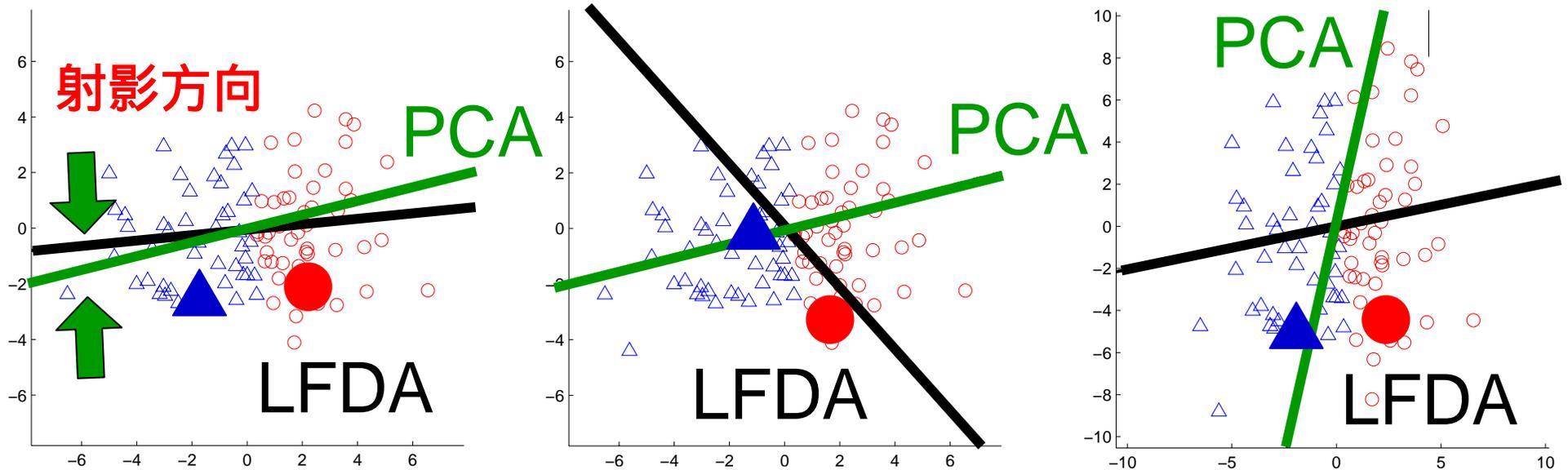
$$\{(x_i, y_i)\}_{i=1}^{n'}$$

- 多数のラベルなし標本

$$\{x_i\}_{i=n'+1}^n$$

- 教師つき次元削減法は, 少数のラベルあり標本に**過適合**しやすい
- 多数のラベルなし標本の情報も活用したい

# 準教師つき学習における LFDAとPCA



- LFDA: 過適合しやすい
- PCA: ラベルの情報を利用していない
- LFDAとPCAは相補的な傾向がある

# 準教師つき局所フィッシャー判別分析<sup>01</sup> (SELF, Semi-Supervised LFDA)

- **基本アイデア**: LFDAとPCAのいいところを組み合わせる
- **着目点**: LFDAとPCAは同じ形式の固有値問題
  - PCA:  $S^{(t)}\varphi = \lambda\varphi$
  - LFDA:  $S^{(lb)}\varphi = \lambda S^{(lw)}\varphi$
- **方針**: 固有値問題を組み合わせる

# SELF (続き)

32

## ■ SELFの固有値問題: LFDAとPCAの重みつき和

$$S^{(rlb)} \varphi = \lambda S^{(rlw)} \varphi$$

- 正則化局所クラス内散布行列

$$S^{(rlw)} = (1 - \beta) S^{(lw)} + \beta I \quad 0 \leq \beta \leq 1$$

- 正則化局所クラス間散布行列

$$S^{(rlb)} = (1 - \beta) S^{(lb)} + \beta S^{(t)}$$

## ■ 解: 上位の固有ベクトルを並べる

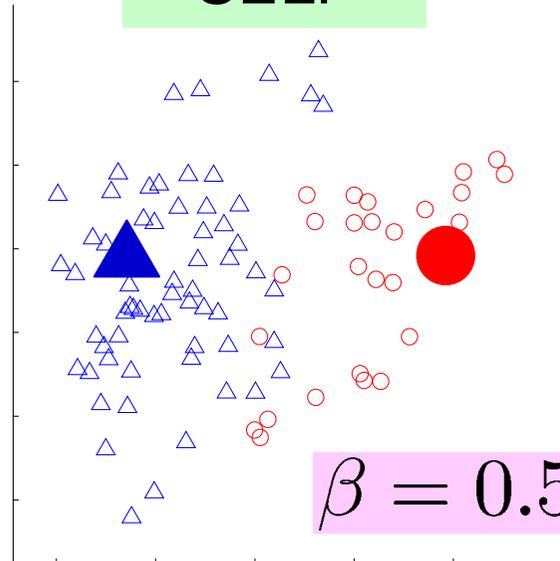
# Olivetti Faceデータの可視化

33

■ メガネ vs. メガネなし

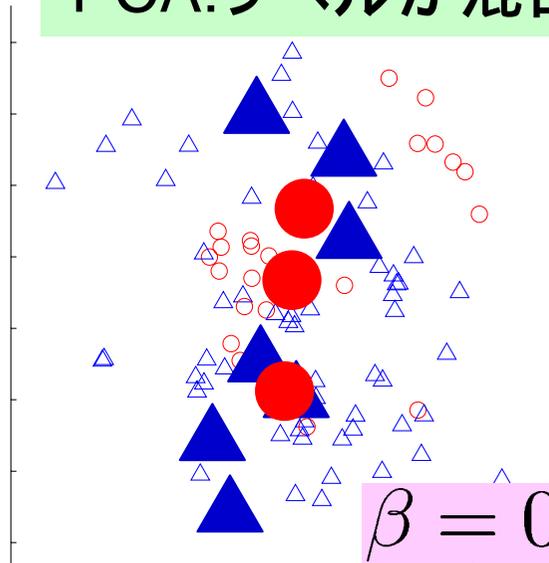


SELF



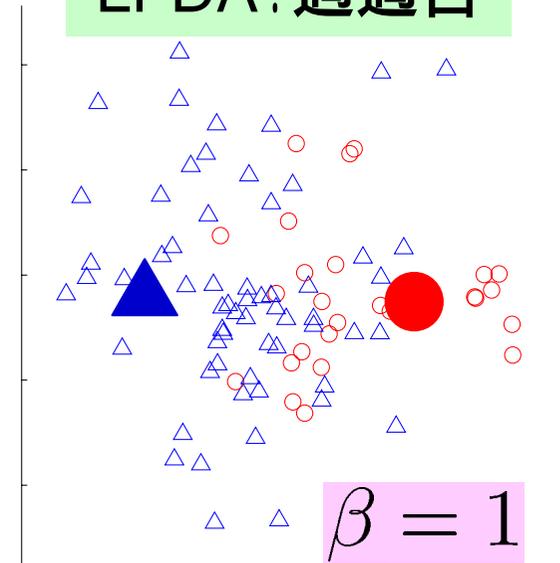
$\beta = 0.5$

PCA:ラベルが混合



$\beta = 0$

LFDA:過適合



$\beta = 1$

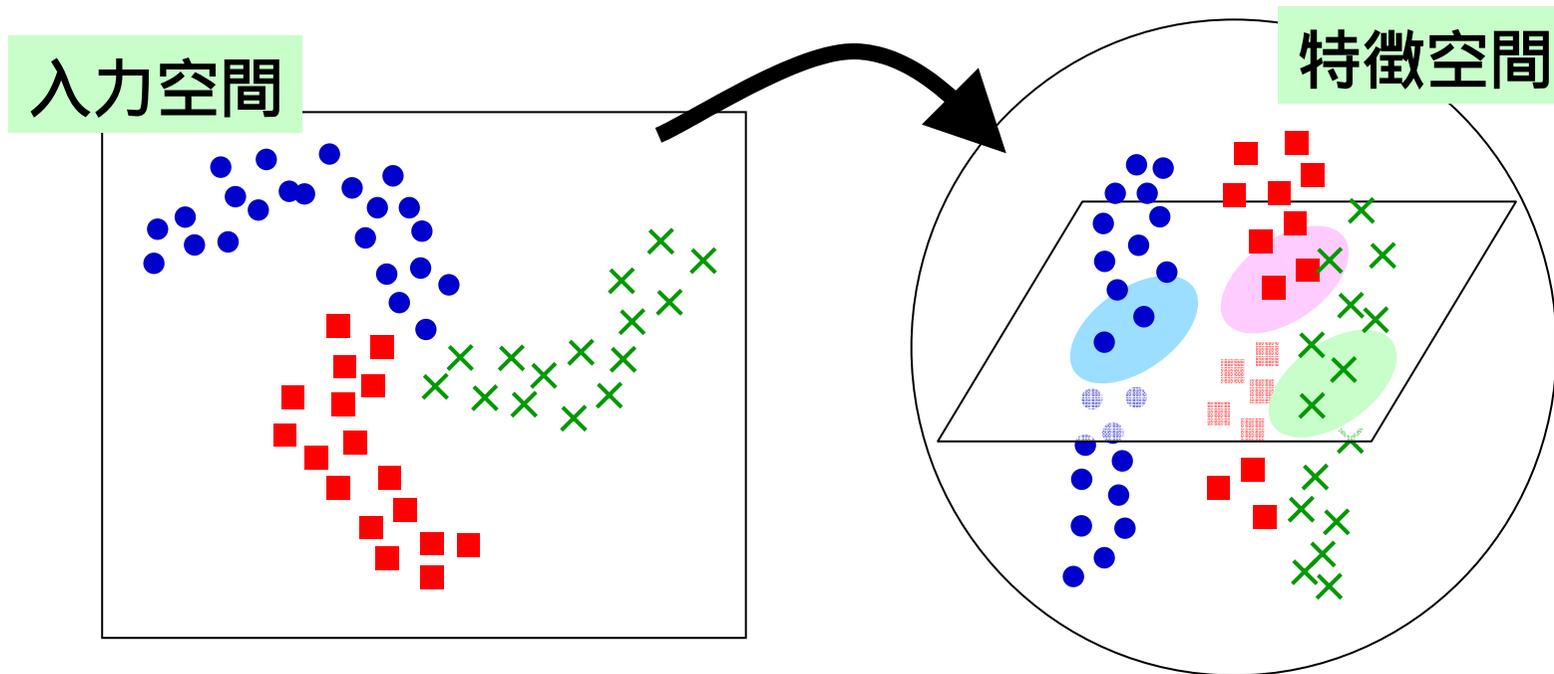
# 次元削減 + 最近傍識別の誤認識率<sup>34</sup>

	LFDA	SELF ( $\beta = 0.5$ )	PCA	SELF (CV)
SSL1	14.9(1.8)	6.0(1.3)	6.2(1.1)	6.0(1.4)
SSL2	15.7(0.9)	9.6(1.1)	11.2(0.8)	10.3(2.4)
SSL3	21.1(3.9)	14.3(1.8)	15.5(1.0)	14.1(1.4)
SSL4	33.4(3.5)	36.6(2.4)	48.7(2.4)	33.4(3.7)
SSL5	27.5(2.3)	27.2(2.3)	31.0(1.9)	27.3(2.9)
SSL6	38.1(1.5)	35.4(2.4)	27.3(2.7)	27.0(2.7)
SSL7	29.4(2.4)	29.1(2.4)	29.3(1.6)	27.7(1.4)

- LFDAとPCAは相補的
- SELF ( $\beta = 0.5$ ) はLFDAとPCAのいいところ取り
- $\beta$  をクロスバリデーションで最適化すると、さらに性能が向上

# 非線形次元削減:カーネルSELF 35

- 標本を**非線形変換**で特徴空間に移す
- 特徴空間内で**線形次元削減**を行なう



- **カーネル関数**を用いることにより, 効率よく非線形次元削減が行なえる

# 本発表の構成

1. 線形次元削減問題
2. 主成分分析(PCA)
3. 局所性保存射影(LPP)
4. フィッシャー判別分析(FDA)
5. 局所フィッシャー判別分析(LFDA)
6. 準教師つき局所フィッシャー判別分析(SELF)
7. 次元削減と計量学習



# 計量学習

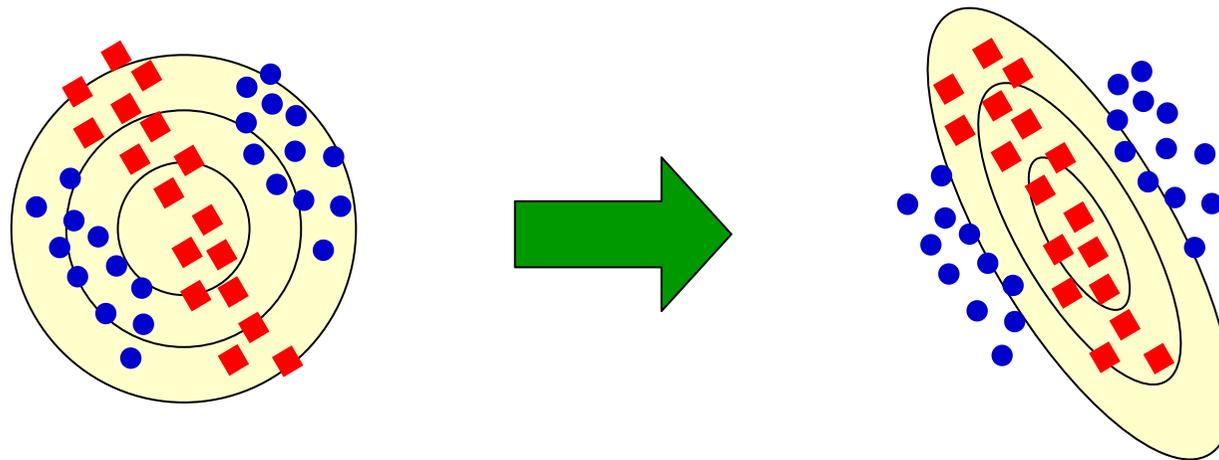
## ■ 計量行列 $M$ :

Xing, Ng, Jordan & Russell (NIPS2002)

Goldberger, Roweis, Hinton & Salakhutdinov (NIPS2004)

Weinberger, Blitzer & Saul (NIPS2005)

$$\text{dist}_M(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top M (\mathbf{x}_i - \mathbf{x}_j)$$



## ■ 計量学習 : 最近傍識別機の性能向上

$$\max_M J(M) \text{ s.t. } M \succeq O$$

## ■ $J(M)$ が凸関数のとき, 大域解を求められる

# 計量学習と次元削減

- $M = TT^\top$  と分解すれば,

$$\text{dist}_M(x_i, x_j) = \|T^\top(x_i - x_j)\|^2$$

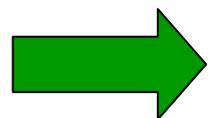
- 計量のランクが縮退しているとき,  
計量学習と同時に  
次元削減も行なう

$$M = T T^\top$$

- しかし, 低ランクの拘束条件は非凸:

$$\text{rank}(M) = r \ (\leq d)$$

- $J(M)$  が凸関数であっても,  $J(TT^\top)$  は  
 $T$  に関しては非凸



次元削減は一般に非凸!

# 次元削減と計量学習 (続き)

39

- レイリー商形式の次元削減:

$$\hat{T} = \operatorname{argmax}_T \left[ \operatorname{tr}(T^\top \bar{C} T (T^\top \underline{C} T)^{-1}) \right]$$

- この最適化問題は凸ではないが, 最適な埋め込み空間の値域が一意に定まる

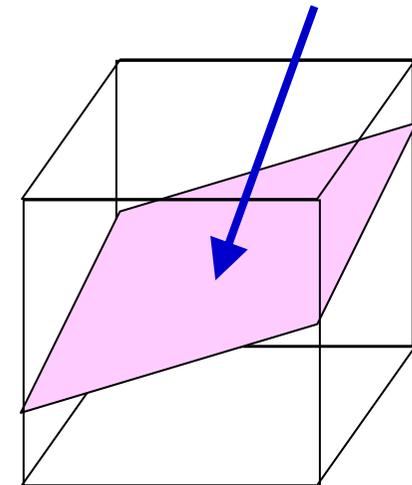
$$\operatorname{range}(\hat{T}) = \operatorname{span}(\{\varphi_k\}_{k=1}^r)$$

(固有空間)

$$\bar{C}\varphi = \lambda \underline{C}\varphi$$

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$$

$\operatorname{range}(\hat{T})$



# 次元削減と計量学習 (続き)

40

$$J(T) = \text{tr}(T^\top \bar{C} T (T^\top \underline{C} T)^{-1})$$

- 但し, レイリー商形式の定式化では, 埋め込み空間内の計量は決定できない:

- $J$  は計量に関して不変

$$J(T) = J(TU)$$

$U$ : 任意の正則行列

- SELFではヒューリスティックに計量を決定:

- 上位の成分の重みを強く
- 下位の成分の重みを弱く

$$T_{SELF} = (\sqrt{\lambda_1} \varphi_1 | \sqrt{\lambda_2} \varphi_2 | \cdots | \sqrt{\lambda_r} \varphi_r)$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$$

# 次元削減と計量学習(続き)

- 固有値による重み付けは, 実験的には良さそうである
- 計量も学習するのであれば, **2段階**で行なうことにより**解を一意に定める**ことができる:

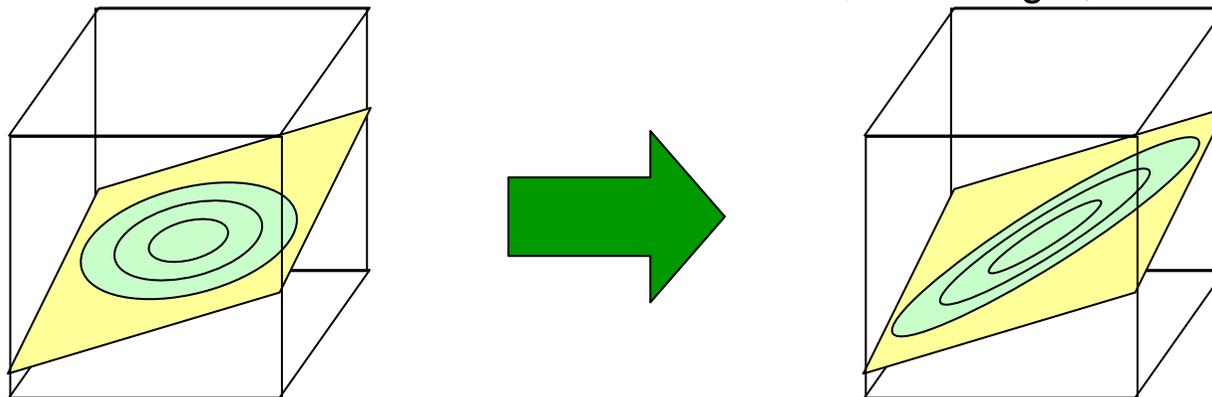
## 第1段階: 次元削減

SELFで埋め込み空間の**値域**を一意に決定

## 第2段階: 計量学習

凸計量学習法で埋め込み空間内の**計量**を一意に決定

例えば, Weinberger, Blitzer & Saul (NIPS2005)



# まとめ



## ■ LFDA: FDAとLPPの組み合わせ

- クラス間分離性とクラス内局所構造を保存
- 解が閉じた形で計算でき, 高速かつ安定

## ■ SELF: LFDAとPCAの組み合わせ

- クラス間分離性, クラス内局所構造, 大域的構造をバランスよく保存
- 解が閉じた形で計算でき, 高速かつ安定

## ■ 計量学習: 次元削減も同時に行なうと凸性が失われ, 解が一意に求まらない

## ■ SELF+凸計量学習の2段階法により, 解を一意に定めることができる

# 謝辞

- 機械学習研究グループ“**T-PRIMAL**”のメンバーのご意見・ご討論に感謝します
- T-PRIMAL (**T**okyo **PR**obabilistic **I**nference and **MA**chine **L**earning)
  - **メンバー**: 杉山将(東工大), 矢入健久(東大), 津田宏治(産総研・MPG), 井手剛(IBM), 鹿島久嗣(IBM), 坪井祐太(IBM), 栗原賢一(東工大), 井上雅史(NII), 加藤毅(東大), 金崎弘文(東大), 河原吉伸(東大), 清水伸幸(東大), 瀬々潤(お茶大), 寺本礼仁(NEC), 中島伸一(ニコン), 浜田道昭(みずほ情報総研), 比戸将平(IBM), 藤巻遼平(NEC), 山崎啓介(東工大), 渡辺一帆(東大), 富岡亮太(東大), 佐藤一誠(東大)
  - **2007年の主要研究業績**: NIPS2007(4本), ICDM2007(2本), SDM2007(1本), ECML2007(1本), ICML2007(3本), JMLR2007(2本), BMC Bioinformatics(1本), Neural Computation(1本)

<http://sugiyama-www.cs.titech.ac.jp/T-PRIMAL>