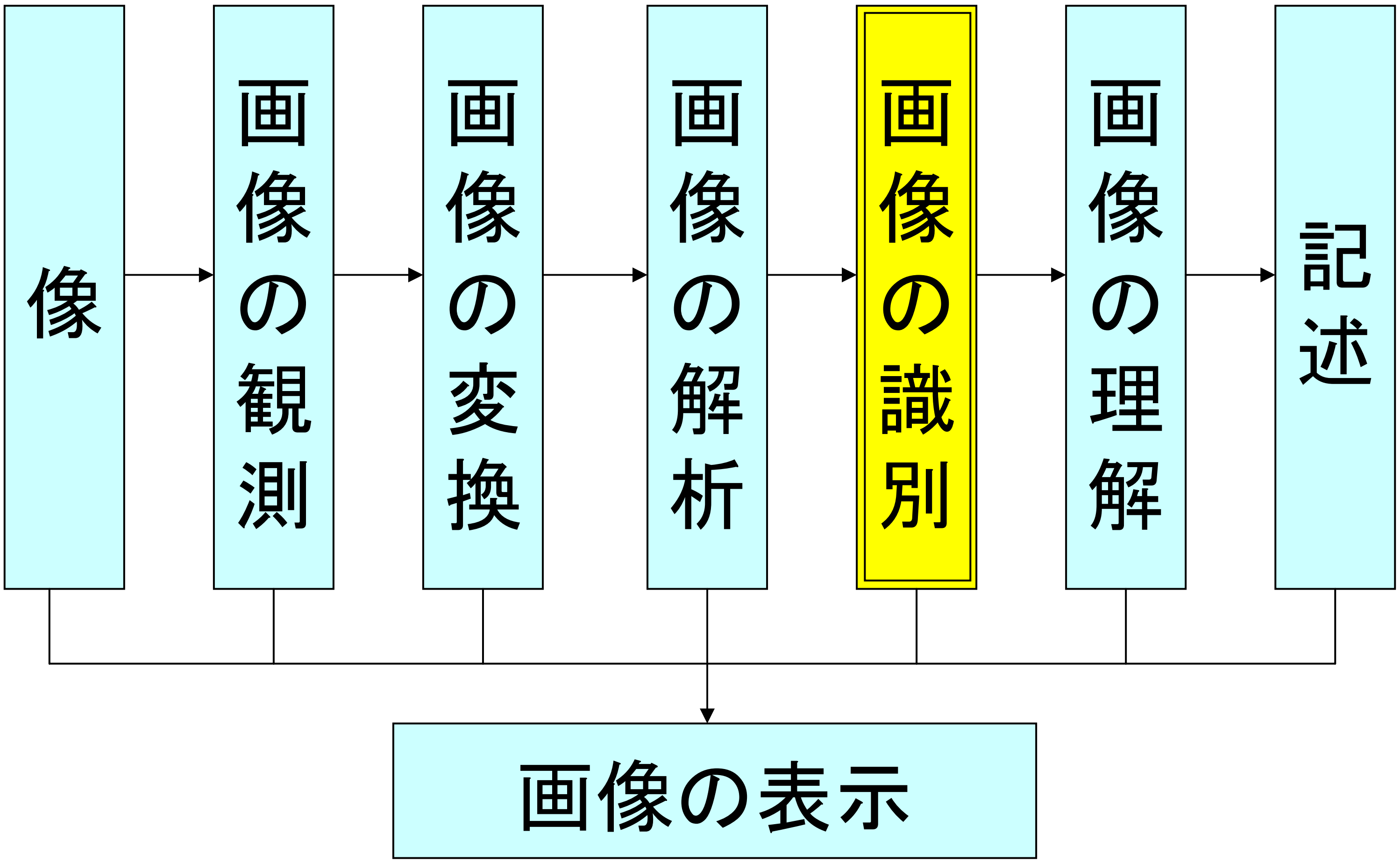
特徴選択と特徴空間

画像の処理工程



画像の識別処理

特徴選択

パターンの識別に用いる特徴を決定

特徴の評価基準の目安

：誤識別率との対応づけが容易であること
：パターン成分に関して線形性があること
：距離測度の条件を満たすこと

⋮

１つの特徴概念で，すべての目安を満たすことは困難．

特徴選択基準(1)

パターン間距離にもとづいた基準

パターン間類似度に基づいた基準

誤識別率に基づいた基準

⋮

選択された特徴 ⇨ 特徴ベクトル

特徴ベクトルの分布 ⇨ 特徴空間

識別の効率（識別率，計算時間等）を考慮した特徴空間の構成が必要

特徴選択基準(2)

距離最小化基準

特徴評価基準の中で最も一般的.

クラス内距離が小さく，クラス間距離が大きくなる特徴を選択すれば識別には有効

距離の種類

ユークリッド距離,

マハラノビス距離

未知パターン $\boldsymbol{x}$,クラスタ C の平均と共分散行列 $\boldsymbol{m}, \boldsymbol{M}$

$d_E^2(\boldsymbol{x},C)=(\boldsymbol{x}-\boldsymbol{m})^T(\boldsymbol{x}-\boldsymbol{m})$：ユークリッド距離

$d_M^2(\boldsymbol{x},C)=(\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{M}^{-1}(\boldsymbol{x}-\boldsymbol{m})$：マハラノビス距離

２つのクラスタC_1, C_2において

　クラスタ中心$\boldsymbol{m}_1, \boldsymbol{m}_2$,

　クラスタ共分散行列$\boldsymbol{M}_1, \boldsymbol{M}_2$,

であるとする.

このとき，ある点 $\boldsymbol{x}$から各クラスタ中心までのユークリッド距離$d_E^2(\boldsymbol{x}, C_1), d_E^2(\boldsymbol{x}, C_2)$が等しい場合でも，クラスタ分散$\boldsymbol{M}_1, \boldsymbol{M}_2$,に応じて

$$d_M^2(\boldsymbol{x}, C_1) \le d_M^2(\boldsymbol{x}, C_2)$$

あるいは,

$$d_M^2(\boldsymbol{x}, C_1) \ge d_M^2(\boldsymbol{x}, C_2)$$

となる.

あるクラスタがn個のd次元列ベクトル $\boldsymbol{x}_i (i = 1, \ldots n)$から構成されているとき，このクラスタの共分散行列$\boldsymbol{M}$は,

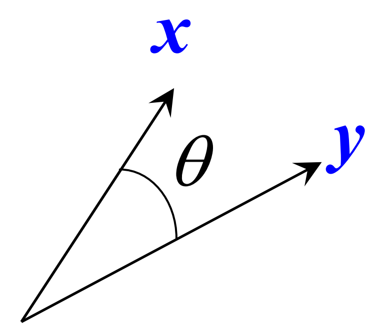
$$\boldsymbol{M} = \frac{1}{n} \sum_{i=1}^{n} \left[(\boldsymbol{x}_i - \hat{\boldsymbol{x}})(\boldsymbol{x}_i - \hat{\boldsymbol{x}})^T \right]$$

特徴選択基準(3)

類似度最大化基準

特徴 $\boldsymbol{x}$ と $\boldsymbol{y}$ の類似度

$$s[\boldsymbol{x},\boldsymbol{y}] = \frac{\boldsymbol{x}^T \boldsymbol{y}}{\|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|} = \cos\theta$$



特徴選択基準

$$D_s = \sum_{i=1}^{N} \sum_{\boldsymbol{x}\in|C_i|} \sum_{\boldsymbol{y}\in|C_i|} s[\boldsymbol{x},\boldsymbol{y}] \quad C_i\,(i=1,\dots,N)$$

が最も大きくなるような特徴を選択.

特徴選択基準(4)

誤識別率最小化基準

準備

特徴空間上のパターンクラス $C_i, (i=1,\cdots,m)$

パターン $\boldsymbol{x}=(x_1, x_2, \dots, x_r)^T$

識別規則
$$\delta(\boldsymbol{x})=\begin{pmatrix}\delta_1(\boldsymbol{x})\\ \vdots \\ \delta_m(\boldsymbol{x})\end{pmatrix}$$

$\delta_i(\boldsymbol{x})$: パターン$\boldsymbol{x}$をクラスC_iに属すると判定する確率

$$\delta_1(\boldsymbol{x})+\delta_2(\boldsymbol{x})+\dots+\delta_m(\boldsymbol{x})=1$$

C_i に帰属するパターンが，δ によって C_j に帰属するとされる確率：

$$P(i \to j;\delta) = E\{\delta_j(x) \mid C_i\} = \int \delta_j(x)\, p(x \mid C_i)\, dx$$

$i \neq j$のとき，C_iのパターンをC_jと誤認識する確率.

$x \in C_i$ が C_i 以外のクラスに誤識別される確率は,

$$P(i;\delta) = \sum_{j=1}^{m} P(i \to j;\delta), \quad j \neq i$$

δ に基づく平均誤識別率は,

$$P_e(\delta) = \sum_{i=1}^{m} P(i;\delta) P(C_i)$$

$P_e(\delta)$を最小にするδ：ベイズの識別規則

（例）

次の非確率的規則δ^*は，ベイズの識別規則である.

$$\delta^*(x) = (\delta_1^*(x), \delta_2^*(x))^T$$

$$\delta_1^*(x) = \begin{cases} 1, & P(C_2)p(x|C_2) \le P(C_1)p(x|C_1) \\ 0, & \text{otherwise} \end{cases},$$

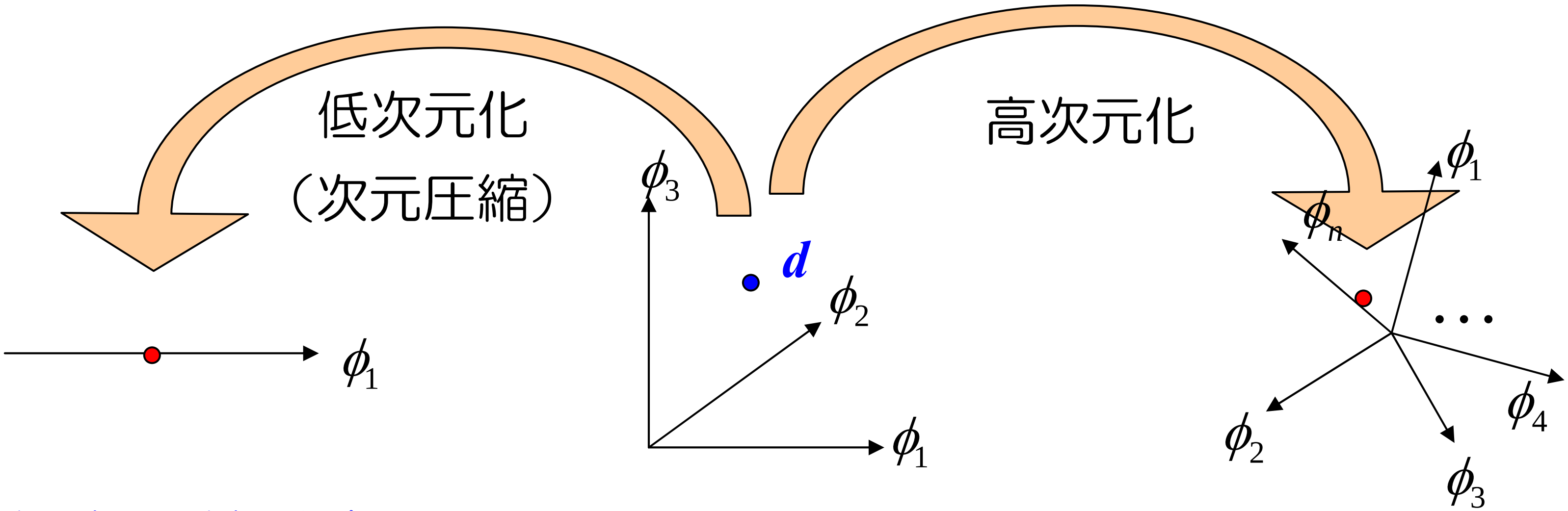
$$\delta_2^*(x) = \begin{cases} 1, & P(C_2)p(x|C_2) > P(C_1)p(x|C_1) \\ 0, & \text{otherwise} \end{cases}$$

この規則は次のようにも表される.

$$\frac{p(x|C_1)}{p(x|C_2)} \begin{cases} \ge P(C_2)/P(C_1) \Rightarrow x \approx C_1 \\ < P(C_2)/P(C_1) \Rightarrow x \approx C_2 \end{cases}$$

特徴空間の次元数

目的に応じて特徴空間の調整が必要



低次元特徴空間
（一次元空間）

原特徴空間

高次元特徴空間
（無限次元空間）

線形分離困難さの増大
分類器の低コスト化

線形分離可能性の増大
分類器の高コスト化

特徴空間の低次元化

特徴数を削減する必要がある場合，その方法としては

d次元原特徴ベクトルから，有用な成分を拾い出し，$\tilde{d}(<d)$次元の新たな特徴ベクトルとして構成する。

ある基準に基づき，d次元原特徴ベクトルを，$\tilde{d}(<d)$次元数の特徴ベクトルに変換する。

→ 特徴空間の変換

多くの場合，

$$\boldsymbol{y} = \boldsymbol{A}^T \boldsymbol{x}, \qquad \boldsymbol{A}: \text{変換行列}$$

と，線形変換として表現．ただし，$\boldsymbol{x}$を原特徴ベクトル（d次元），$\boldsymbol{y}$を変換後の特徴ベクトル（$\tilde{d}$次元）とする．

特徴空間の変換と次元削減(1)

特徴空間の次元削減の基準として，分散最大化基準と平均二乗誤差最小化基準が一般的に用いられる。

いずれの基準に対しても，ＫＬ展開（主成分分析）による特徴空間の変換と，それに基づく次元削減が行われる。

分散最大化基準：

変換後の部分空間においてパターン分布の分散が最大になるように変換する。

平均二乗誤差最小化基準：

変換前後の誤差の平均二乗誤差が最小になるように，変換する

分散最大基準による次元削減

特徴ベクトル集合 $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\},\ \boldsymbol{x}_i \in R^d$

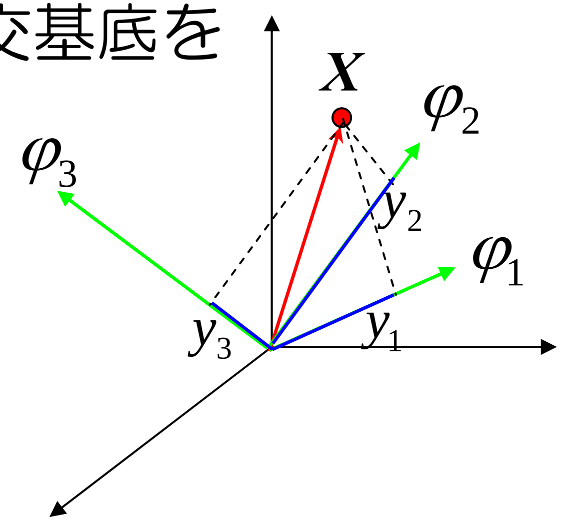
線形変換後の特徴ベクトル集合 $\boldsymbol{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n\},\ \boldsymbol{y}_i \in R^{\tilde{d}}$

変換後の$\tilde{d}\,(< d)$次元部分空間を張る正規直交基底を

$$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_{\tilde{d}}\},\quad \boldsymbol{u}_i \in R^d, (i = 1, \cdots, \tilde{d})$$

とすると，以下の関係が成り立つ．

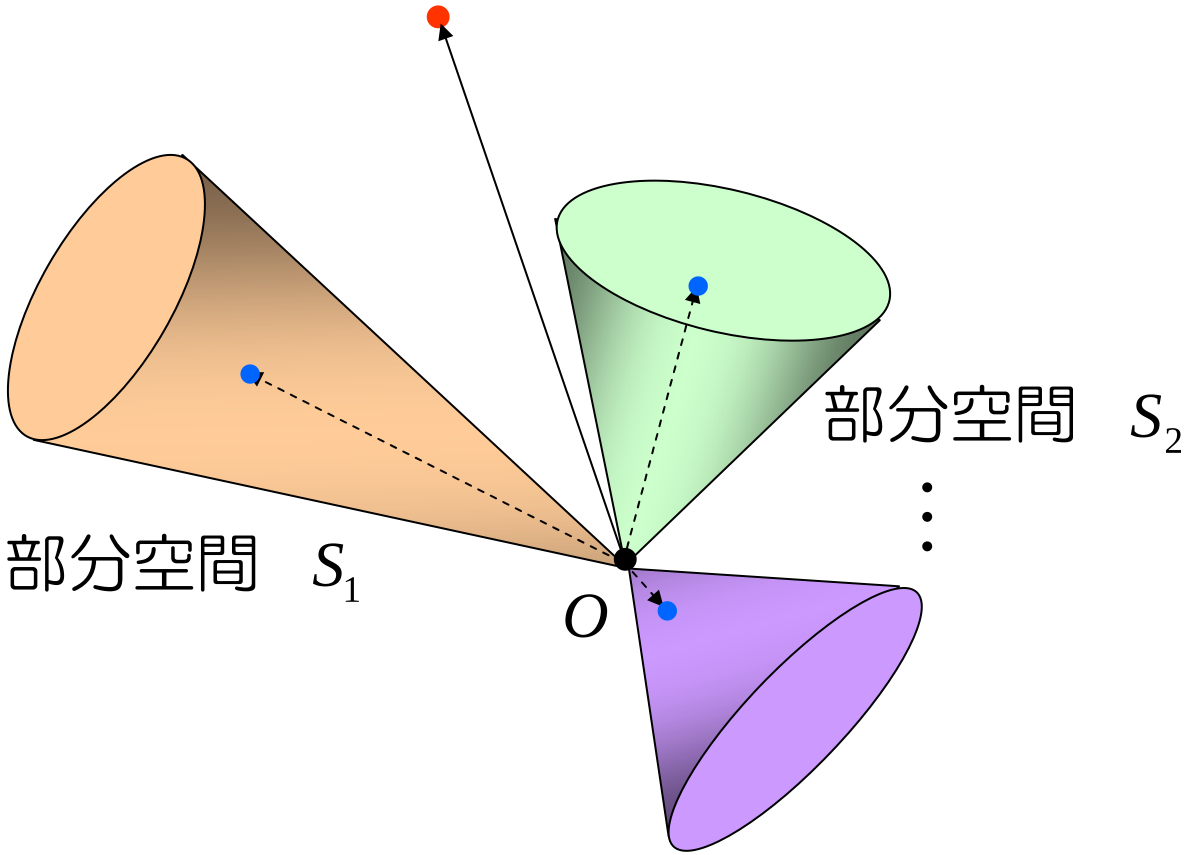
$$\boldsymbol{u}_i \boldsymbol{u}_j = \begin{cases} 1, & i = j \text{のとき} \\ 0, & i \neq j \text{のとき} \end{cases}$$



元の特徴空間から部分空間への変換行列 $\boldsymbol{A}$ は，

$$\boldsymbol{A} = \left(\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_{\tilde{d}}\right),\quad \boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{I}$$

である．ただし，Iは$\tilde{d}$次元単位行列．


部分空間 S_2
部分空間 S_1
O

$$\boldsymbol{m} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i,$$ 原特徴空間でのパターン平均,

$$\tilde{\boldsymbol{m}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i,$$ 部分空間での平均,

とすると,

$$\tilde{\boldsymbol{m}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}^T\boldsymbol{x}_i = \boldsymbol{A}^T\boldsymbol{m}$$

従って，変換行列$\boldsymbol{A}$による部分空間での分散$\sigma(\boldsymbol{A})$は

$$\sigma(A) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \tilde{\boldsymbol{m}})^T(\boldsymbol{y}_i - \tilde{\boldsymbol{m}}) = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{A}^T(\boldsymbol{x}_i - \boldsymbol{m})\right)^T\left(\boldsymbol{A}^T(\boldsymbol{x}_i - \boldsymbol{m})\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} tr\left(\left(\boldsymbol{A}^T(\boldsymbol{x}_i - \boldsymbol{m})\right)\left(\boldsymbol{A}^T(\boldsymbol{x}_i - \boldsymbol{m})\right)^T\right)$$

$$= tr\left(\boldsymbol{A}^T\frac{1}{n}\sum_{i=1}^{n}\left((\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i - \boldsymbol{m})^T\right)\boldsymbol{A}\right) = tr\left(\boldsymbol{A}^T\boldsymbol{\Sigma}\boldsymbol{A}\right)$$

前式の$\boldsymbol{\Sigma}$は，原特徴空間でのパターン集合の共分散行列で

$$\boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i - \boldsymbol{m})^T$$

分散を最大にする変換行列 $\boldsymbol{A}$ を求める問題

$$\max_{\boldsymbol{A}} : \quad tr\left(\boldsymbol{A}^T \boldsymbol{\Sigma} \boldsymbol{A}\right)$$

$$\text{subj.to} \quad \boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{I}$$

制約付最適化問題

$L(\boldsymbol{A},\lambda) = tr\left(\boldsymbol{A}^T \boldsymbol{\Sigma} \boldsymbol{A}\right) + \lambda(\boldsymbol{I} - \boldsymbol{A}^T \boldsymbol{A});$　ラグランジュ関数

$$\frac{\partial L(\boldsymbol{A},\lambda)}{\partial \boldsymbol{A}} = \boldsymbol{0} \quad \Rightarrow 2\boldsymbol{\Sigma}\boldsymbol{A} - 2\lambda\boldsymbol{A} = 2(\boldsymbol{\Sigma} - \lambda\boldsymbol{I})(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{\tilde{d}}) = \boldsymbol{0}$$

$$\frac{\partial L(\boldsymbol{A},\lambda)}{\partial \lambda} = \boldsymbol{0} \quad \Rightarrow \boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{I}$$

最適化問題の解 ⇨ $\det(\boldsymbol{\Sigma} - \lambda\boldsymbol{I}) = 0$（固有方程式）の解と等価．

従って，共分散行列$\boldsymbol{\Sigma}$の固有値$\lambda_1,\cdots,\lambda_d\,(\lambda_1 \geq \cdots \geq \lambda_d)$と，
対応する固有ベクトル$\boldsymbol{u}_1,\cdots,\boldsymbol{u}_d$は，$\boldsymbol{\Sigma}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$　と表される．

一方，$\boldsymbol{\Sigma}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$の関係を行列表現すると

$$\boldsymbol{\Sigma}(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_d) = (\lambda_1\boldsymbol{u}_1,\cdots,\lambda_d\boldsymbol{u}_d) \quad \Rightarrow \quad \boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{\Lambda}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0\cdots & 0 \\ 0 & \lambda_2\cdots & 0 \\ & \cdots & \\ 0 & 0\cdots & \lambda_d \end{pmatrix}$$

と表すことが可能．従って，

$$\boldsymbol{A}^T\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{\Lambda} \quad (\because \boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I})$$

これより，変換後の分散$\sigma^2(\boldsymbol{A}) = tr\left(\boldsymbol{A}^T\boldsymbol{\Sigma}\boldsymbol{A}\right)$の最大値は

$$\max\left\{\sigma^2(\boldsymbol{A})\right\} = \max\left\{tr(\boldsymbol{\Lambda})\right\} = \sum_{i=1}^{\tilde{d}} \lambda_i$$

共分散行列$\boldsymbol{\Sigma}$の大きい$\tilde{d}$個の固有値に対応する固有ベクトル
$\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{\tilde{d}}$からなる行列$\boldsymbol{A}$での変換が分散最大化を達成する．

平均二乗誤差最小基準次元削減

変換行列Aによって変換した特徴ベクトルを，再度元の特徴空間から眺めたときの，原特徴ベクトルとの平均二乗誤差を最小化する

特徴ベクトル集合 $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\},\ \boldsymbol{x}_i \in R^d$

線形変換後の特徴ベクトル集合 $\boldsymbol{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n\},\ \boldsymbol{y}_i \in R^{\tilde{d}}$

変換行列$\boldsymbol{A}, (d \times \tilde{d})$による変換：$\boldsymbol{y}_i = \boldsymbol{A}^T \boldsymbol{x}_i$

特徴ベクトル $\boldsymbol{y}_i$ を，原空間で展開した点を $\tilde{\boldsymbol{x}}_i$ とすると，

$$
\begin{aligned}
\tilde{\boldsymbol{x}}_i &= y_{i1}\boldsymbol{u}_1 + y_{i2}\boldsymbol{u}_2 + \cdots + y_{i\tilde{d}}\boldsymbol{u}_{\tilde{d}} \\
&= (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{\tilde{d}})(y_{i1}, \cdots, y_{i\tilde{d}})^T = \boldsymbol{A}\boldsymbol{y}_i
\end{aligned}
$$

変換Aによる平均二乗誤差を$\varepsilon^2(A)$とすると，

$$\varepsilon^2(A) = E\left\{\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|^2\right\} = \frac{1}{n}\sum_{i=1}^{n}\left\{(A\boldsymbol{y}_i - \boldsymbol{x}_i)^T(A\boldsymbol{y}_i - \boldsymbol{x}_i)\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{(AA^T\boldsymbol{x}_i - \boldsymbol{x}_i)^T(AA^T\boldsymbol{x}_i - \boldsymbol{x}_i)\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{x}_i^T\boldsymbol{x}_i - (A^T\boldsymbol{x}_i)^T A^T\boldsymbol{x}_i\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{tr(\boldsymbol{x}_i\boldsymbol{x}_i^T) - tr((A^T\boldsymbol{x}_i\boldsymbol{x}_i^T A)^T)\right\}$$

$$= tr(\boldsymbol{R}) - tr(A^T\boldsymbol{R}A)$$

ここで，$\boldsymbol{R} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T$　：　自己相関行列．

上記の平均二乗誤差の最小化問題

⟹ $tr(A^T\boldsymbol{R}A)$を最大化する問題

自己相関行列$\boldsymbol{R}$と共分散行列$\boldsymbol{\Sigma}$の関係

$$\boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i - \boldsymbol{m})^T = \boldsymbol{R} - \boldsymbol{m}\boldsymbol{m}^T$$

$\boldsymbol{R}$の固有値：　$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$

対応する固有ベクトル：　$\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_d$

このとき，$\tilde{d}$個の$\lambda_i, (i = 1, \cdots, \tilde{d})$に対応する固有ベクトル行列

$$\boldsymbol{A} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{\tilde{d}})$$

は，平均二乗誤差最小基準を満たす変換行列となる．

最小誤差は

$$\min\{\varepsilon^2(\boldsymbol{A})\} = tr(\boldsymbol{R}) - \sum_{i=1}^{\tilde{d}}\lambda_i = \sum_{i=\tilde{d}}^{d}\lambda_i$$

特徴空間の変換と次元削減(2)

特徴空間の次元削減のもう１つの代表的基準として，線形判別基準（判別分析基準）がある。

この基準は：

特徴空間と各特徴のクラスが既知である場合，クラス内分散とクラス間分散の比を最大にするというもの。

線形判別法(Linear Discriminant method)による変換：

ＫＬ展開がパターン全体の分布の最良近似を目的としているのに対し，線形判別は識別のためのクラス毎のパターン分離度を最大化することを目的とする。

特に，２クラスの分類を対象とした方法を，
フィッシャーの方法(Fisher’s method)と呼ぶ。

フィッシャーの線形判別法

d 次元特徴ベクトル$\boldsymbol{x} \in R^d$のクラス$\omega_i, (i=1,2)$の変動行列(scatter matrix)を$\boldsymbol{S}_i$，平均ベクトルを$\boldsymbol{m}_i$とすると，

$$\boldsymbol{S}_i = \sum_{\boldsymbol{x} \in \omega_i} (\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^T, \boldsymbol{x} \in R^d$$

$$\boldsymbol{m}_i = \frac{1}{n_i} \sum_{\boldsymbol{x} \in \omega_i} \boldsymbol{x}, \quad n_i \text{は} \omega_i \text{のパターン数}$$

さらに，クラス内変動行列(within - class scatter matrix)$\boldsymbol{S}_W$とクラス間変動行列(between - class scatter matrix)$\boldsymbol{S}_B$を，

$$\boldsymbol{S}_W = \boldsymbol{S}_1 + \boldsymbol{S}_2 = \sum_{i=1,2} \sum_{x \in \omega_i} (\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^T,$$

$$\boldsymbol{S}_B = \sum_{i=1,2} n_i (\boldsymbol{m}_i - \boldsymbol{m})(\boldsymbol{m}_i - \boldsymbol{m})^T = \frac{n_1 n_2}{n} (\boldsymbol{m}_1 - \boldsymbol{m}_2)(\boldsymbol{m}_1 - \boldsymbol{m}_2)^T,$$

$$n = n_1 + n_2$$

フィッシャー法では，クラス間変動のクラス内変動に対する比を最大化するように特徴空間を変換する．

変換行列を$\boldsymbol{A}$とすると，

変換後のクラス内変動行列，クラス間変動行列$\tilde{\boldsymbol{S}}_W, \tilde{\boldsymbol{S}}_B$は，

$$\omega_1 : \boldsymbol{m}_1, \boldsymbol{S}_1$$

$$\omega_2 : \boldsymbol{m}_2, \boldsymbol{S}_2$$

$$\boldsymbol{S}_W, \boldsymbol{S}_B$$

$$\boldsymbol{y} = \boldsymbol{A}^T \boldsymbol{x}$$

$$\tilde{\boldsymbol{S}}_W = \tilde{\boldsymbol{S}}_1 + \tilde{\boldsymbol{S}}_1 = \boldsymbol{A}^T \boldsymbol{S}_W \boldsymbol{A}$$

$$\tilde{\boldsymbol{S}}_B = \boldsymbol{A}^T \boldsymbol{S}_B \boldsymbol{A}$$

$$\tilde{\boldsymbol{m}}_i = \frac{1}{n_i} \sum_{x \in X} \boldsymbol{A}^T \boldsymbol{x} = \boldsymbol{A}^T \boldsymbol{m}_i$$

$$J_S(A) = \frac{\tilde{\boldsymbol{S}}_B}{\tilde{\boldsymbol{S}}_W} = \frac{\boldsymbol{A}^T \boldsymbol{S}_B \boldsymbol{A}}{\boldsymbol{A}^T \boldsymbol{S}_W \boldsymbol{A}}$$の最大化

J_Sの最大化問題：

制約条件：$\tilde{\boldsymbol{S}}_W = \boldsymbol{A}^T \boldsymbol{S}_W \boldsymbol{A} = \boldsymbol{I}$の下で
$\tilde{\boldsymbol{S}}_B = \boldsymbol{A}^T \boldsymbol{S}_B \boldsymbol{A}$を最大化する問題

$$J(\boldsymbol{A}) = \boldsymbol{A}^T \boldsymbol{S}_B \boldsymbol{A} - \lambda(\boldsymbol{A}^T \boldsymbol{S}_W \boldsymbol{A} - \boldsymbol{I})$$

をAで変分して0とおく。

$$(\boldsymbol{S}_B + \boldsymbol{S}_B{}^T)\boldsymbol{A} - \lambda(\boldsymbol{S}_W + \boldsymbol{S}_W{}^T)\boldsymbol{A} = \boldsymbol{0}$$

S_B, S_Wは対称行列

$$\boldsymbol{S}_B \boldsymbol{A} - \lambda \boldsymbol{S}_W \boldsymbol{A} = \boldsymbol{0}$$

$$(\boldsymbol{S}_W{}^{-1} \boldsymbol{S}_B - \lambda \boldsymbol{I})\boldsymbol{A} = \boldsymbol{0} \quad (\boldsymbol{S}_W\text{を正則とする})$$

$S_W{}^{-1} S_B$は,(d, d)次元の正方行列．この 行列から求まる固有値 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$の中で，２クラス分類 問題で非零となるのは λ_1のみ.

変換行列$\boldsymbol{A}$は，固有値λ_1に対応する固有ベクトル$\in R^d$.

d次元特徴ベクトルのクラス$\omega_i, (i = 1, \cdots, c)$に対する線形判別法では，$\tilde{d} (\leq c - 1)$次元の特徴空間へ変換される.

クラス内共分散行列を$\boldsymbol{\Sigma}_W$，クラス間共分散行列を$\boldsymbol{\Sigma}_B$とするとき，行列$\boldsymbol{\Sigma}_W{}^{-1}\boldsymbol{\Sigma}_B$の固有値$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{\tilde{d}}$に対応する固有ベクトル列を変換行列$\boldsymbol{A}$とする.

多クラスへ拡張した線形判別法：

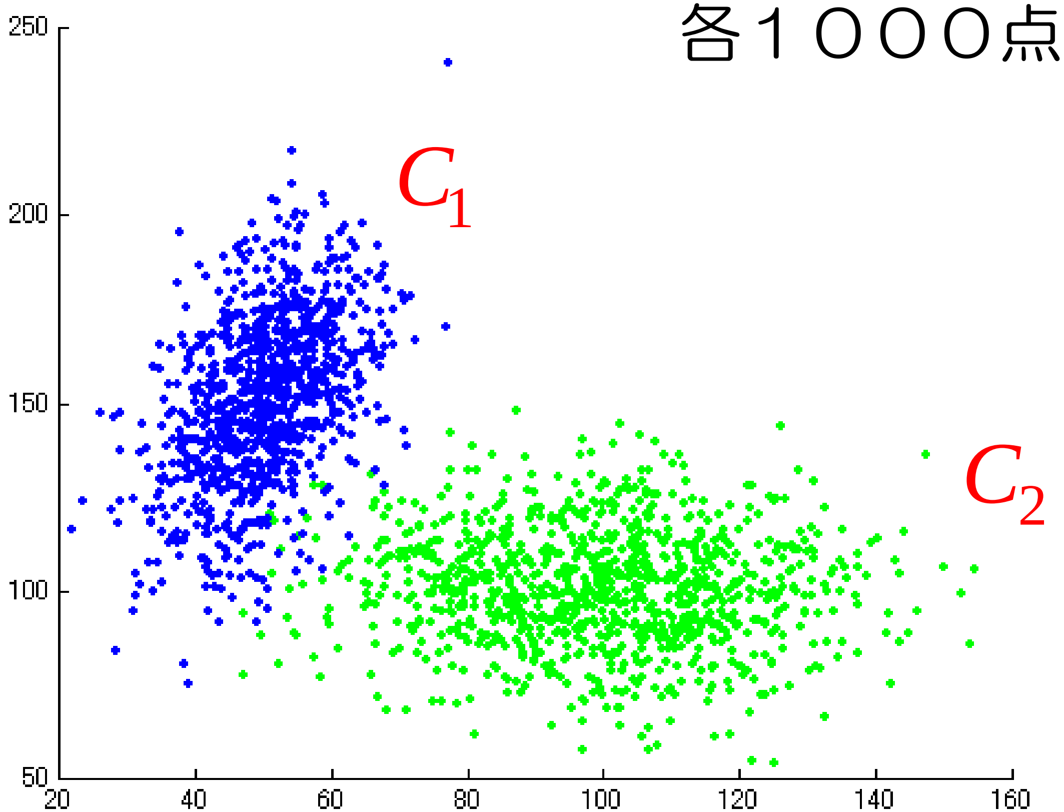
正準判別法(canonical discriminant method)

重判別法(multiple discriminant method)

と呼ばれる.

特徴空間の比較

（例）２つの異なる２次元正規分布
各１０００点



分散最大化基準に基づく特徴変換

$$\boldsymbol{m} = \frac{1}{n_1 + n_2} \sum_{\boldsymbol{x} \in C_1 + C_2} \boldsymbol{x}$$

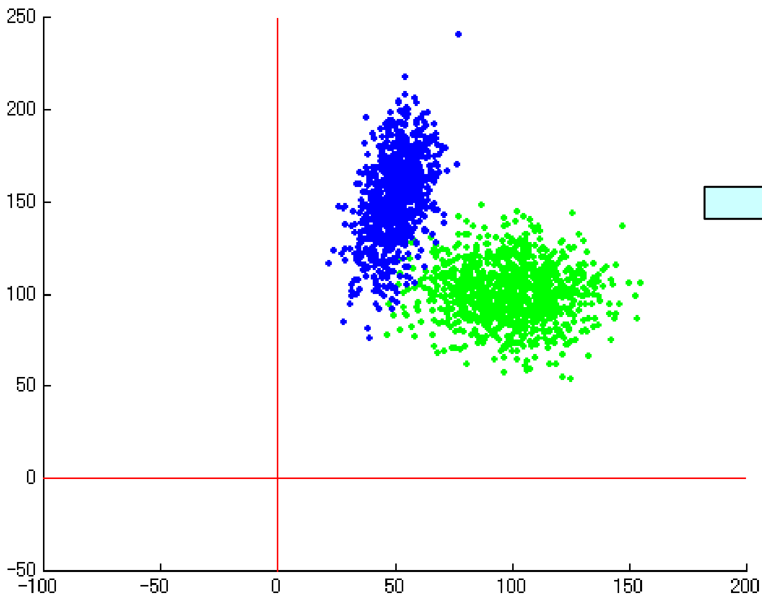
$$\boldsymbol{\Sigma}_X = E\{(\boldsymbol{x} - \boldsymbol{m})(\boldsymbol{x} - \boldsymbol{m})^T\}$$

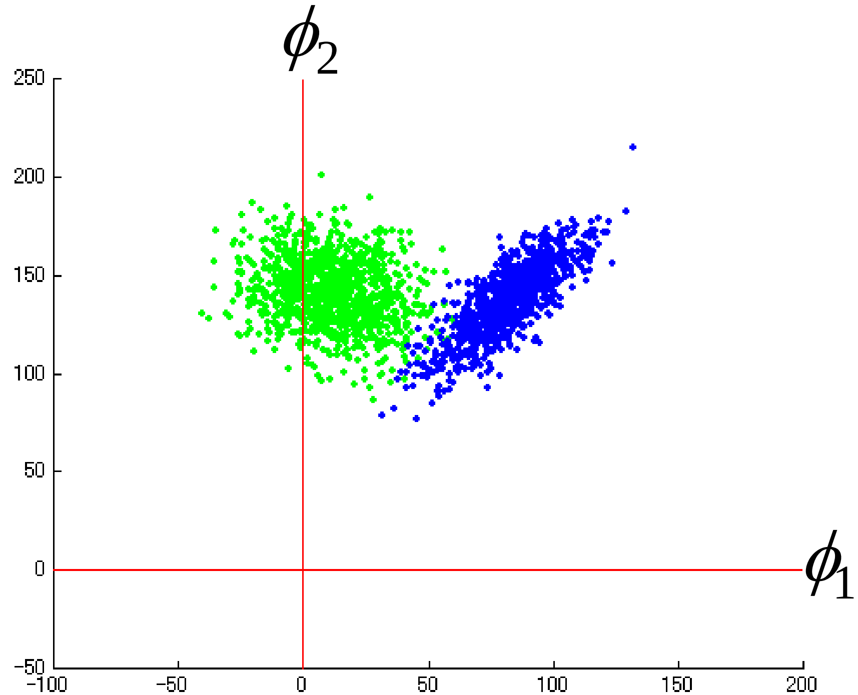
$$= \frac{1}{n_1 + n_2} \sum_{x \in C_1 + C_2} (\boldsymbol{x} - \boldsymbol{m})(\boldsymbol{x} - \boldsymbol{m})^T$$

$$(\boldsymbol{\Sigma}_X - \lambda \boldsymbol{I})\phi_i = \boldsymbol{0}$$

$$\boldsymbol{\Phi} = (\phi_1, \phi_2)$$

変換式：$\boldsymbol{y} = \boldsymbol{\Phi}^T \boldsymbol{x}$





線形判別分析基準に基づく特徴変換

$$\omega_1 : \boldsymbol{m}_1, \boldsymbol{S}_1$$

$$\omega_2 : \boldsymbol{m}_2, \boldsymbol{S}_2$$

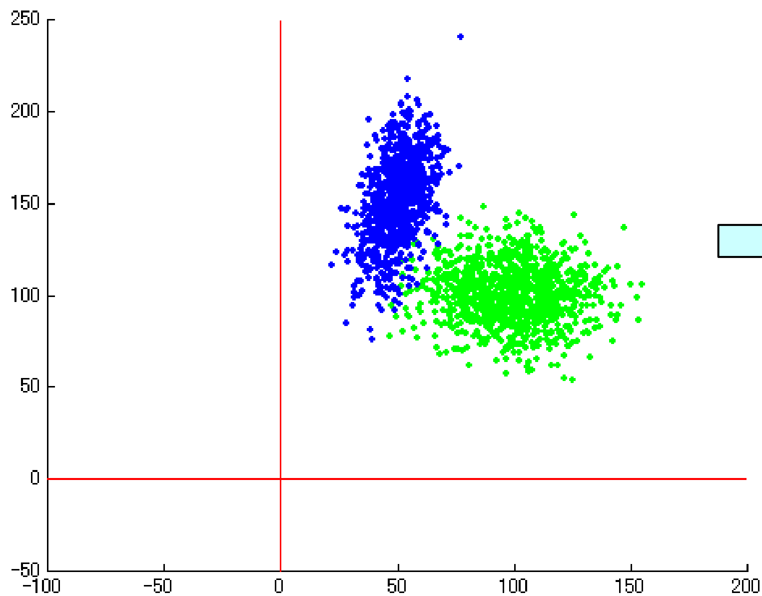
$$\boldsymbol{S}_W = \sum_{i=1,2} \sum_{\boldsymbol{x} \in C_i} (\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^T$$

$$\boldsymbol{S}_B = \frac{n_1 n_2}{n} (\boldsymbol{m}_1 - \boldsymbol{m}_2)(\boldsymbol{m}_1 - \boldsymbol{m}_2)^T$$

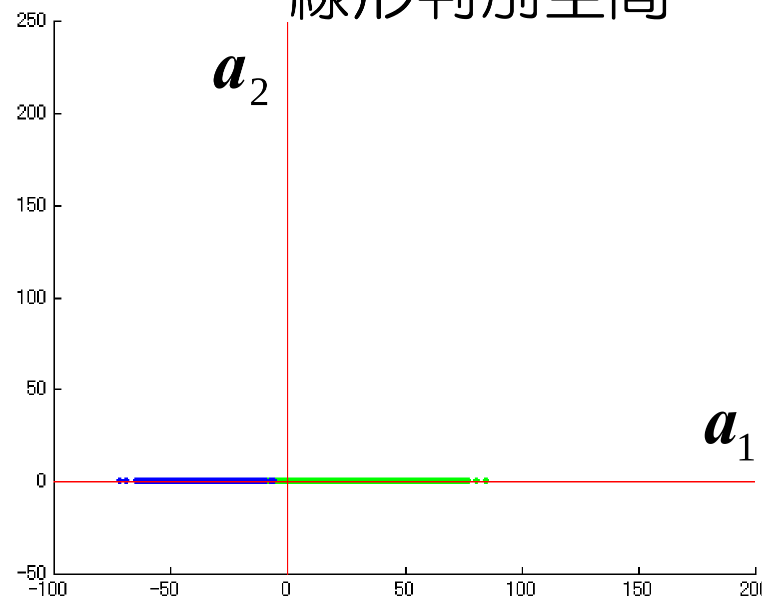
$$(\boldsymbol{S}_W^{-1} \boldsymbol{S}_B - \lambda \boldsymbol{I}) \boldsymbol{a}_i = \boldsymbol{0}$$

$$\boldsymbol{A} = (\boldsymbol{a}_1)$$

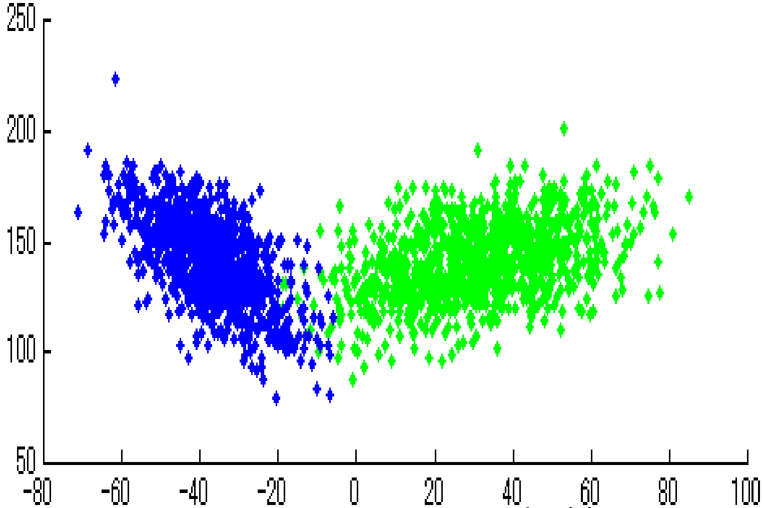
変換式：$\boldsymbol{y} = \boldsymbol{A}^T \boldsymbol{x}$



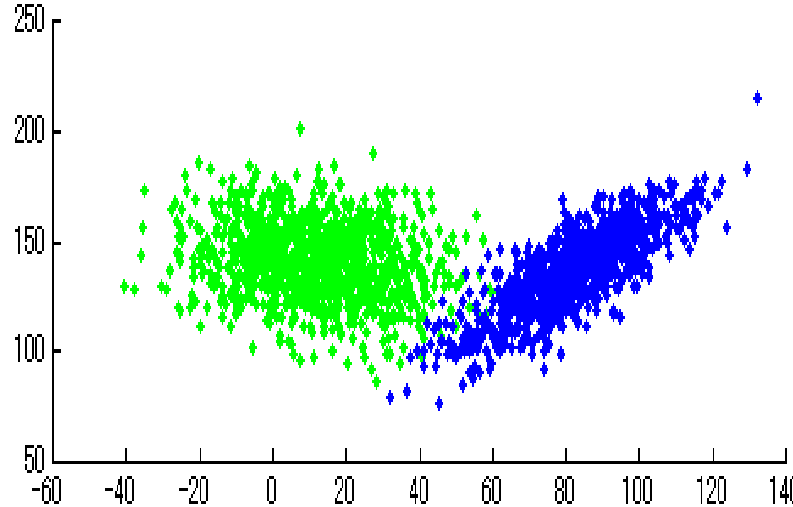
線形判別空間



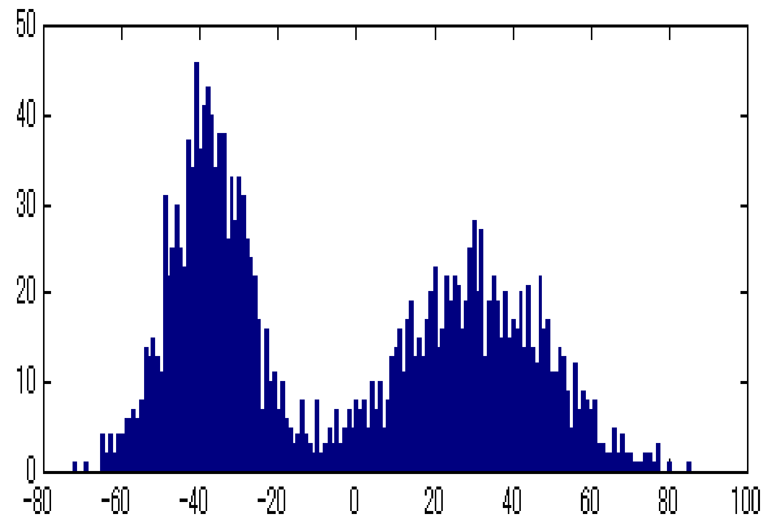
特徴変換後の第一成分軸への射影



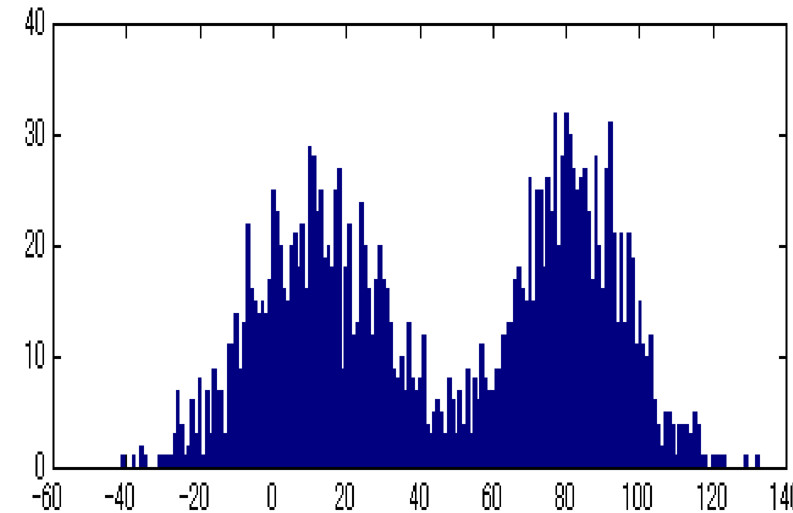
LDAによる変換



PCAによる変換



ヒストグラム



ヒストグラム

特徴空間の高次元化

特徴ベクトルが非線形な構造を有する場合，その解析は困難．

⇒ 特徴ベクトルの高次元空間への非線形写像
非線形な相関解析が可能

（例）２次元特徴ベクトル：　$\boldsymbol{x} = (x_1, x_2)^T$

２次式による非線形写像（５次元特徴化）

$$\boldsymbol{x} = (x_1, x_2)^T \quad \Rightarrow \quad \boldsymbol{X} = (x_1, x_2, {x_1}^2, x_1 x_2, {x_2}^2)$$

一般に，

原空間：m次元 ⇒ d次式の非線形写像

写像空間の次元数：$\dfrac{(m+d-1)!}{d!(m-1)!}$　　膨大な計算量が必要