

パターン認識 —判別分析—

担当: 和田 俊和
部屋 A513

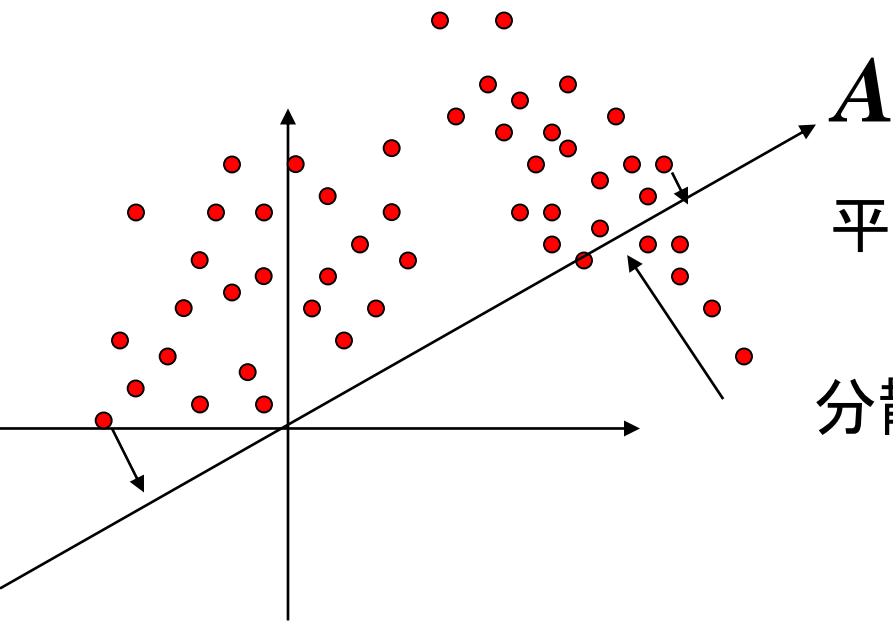
Email twada@ieee.org

講義資料は<http://vrl.sys.wakayama-u.ac.jp/PRA/WPR-2>

判別分析と統計的識別

射影特徴の平均と分散

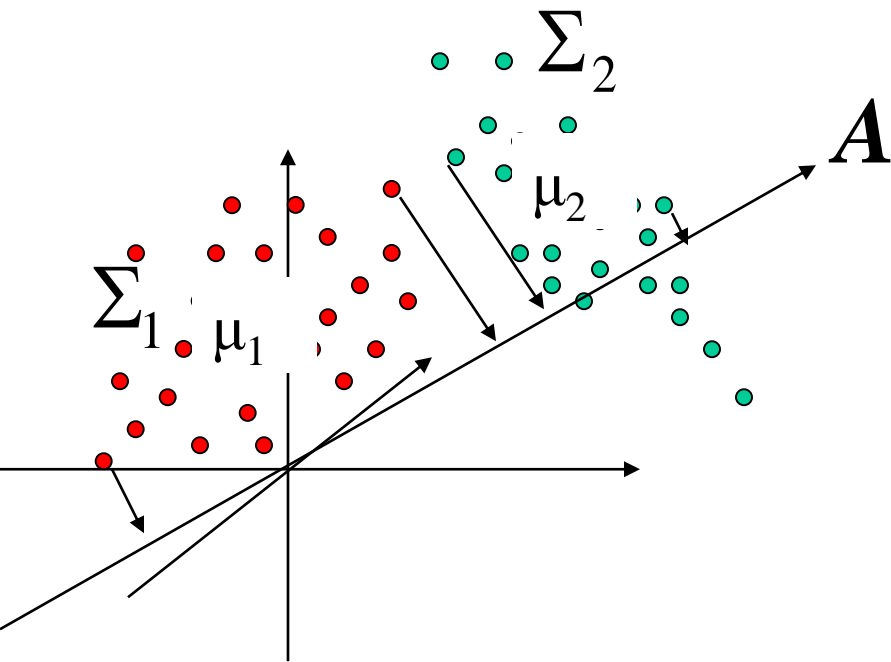
射影後の位置 $y = A^T \mathbf{x}$



平均値 $\frac{1}{n} \sum_{i=1}^n A^T \mathbf{x}_i = A^T \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \boxed{A^T \bar{\mathbf{x}}}$

分散 $\frac{1}{n} \sum_{i=1}^n \left\{ (\mathbf{x}_i - \bar{\mathbf{x}})^T A \right\}^T \left\{ (\mathbf{x}_i - \bar{\mathbf{x}})^T A \right\}$
 $= \frac{1}{n} \sum_{i=1}^n A^T (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T A$
 $= \boxed{A^T \Sigma A}$

クラス内の分散



	クラス1	クラス2
平均値	$\mathbf{A}^T \bar{\mathbf{x}}_1$	$\mathbf{A}^T \bar{\mathbf{x}}_2$
分散	$\mathbf{A}^T \Sigma_1 \mathbf{A}$	$\mathbf{A}^T \Sigma_2 \mathbf{A}$

クラス内分散の期待値

$$\begin{aligned} & \frac{1}{n_1 + n_2} (n_1 \mathbf{A}^T \Sigma_1 \mathbf{A} + n_2 \mathbf{A}^T \Sigma_2 \mathbf{A}) \\ &= \mathbf{A}^T \frac{n_1 \Sigma_1 + n_2 \Sigma_2}{n_1 + n_2} \mathbf{A} \\ &= \boxed{\mathbf{A}^T \Sigma_W \mathbf{A}} \end{aligned}$$

クラス間分散


$$\Sigma_B = \frac{n_1 n_2 (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T}{(n_1 + n_2)^2}$$

$$\text{全分散} V = \frac{1}{n} \sum_{i=1}^n \left((\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{A} \right)^T \left((\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{A} \right)$$

$$= \frac{n_1}{n_1 + n_2} \frac{1}{n_1} \sum_{i=1}^{n_1} \left((\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{A} \right)^T \left((\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{A} \right)$$

$$+ \frac{n_2}{n_1 + n_2} \frac{1}{n_2} \sum_{i=n_1+1}^n \left((\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{A} \right)^T \left((\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{A} \right)$$

$$\bar{\mathbf{x}} = \frac{n_1 \bar{\mathbf{x}}_1 + n_2 \bar{\mathbf{x}}_2}{n_1 + n_2}$$



$$= \frac{n_1}{n_1 + n_2} \frac{1}{n_1} \sum_{i=1}^{n_1} \left(\left(\mathbf{x}_i - \bar{\mathbf{x}}_1 + \frac{n_2(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{n_1 + n_2} \right)^T \mathbf{A} \right)^T \left(\left(\mathbf{x}_i - \bar{\mathbf{x}}_1 + \frac{n_2(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{n_1 + n_2} \right)^T \mathbf{A} \right)$$

$$+ \frac{n_2}{n_1 + n_2} \frac{1}{n_2} \sum_{i=n_1+1}^n \left(\left(\mathbf{x}_i - \bar{\mathbf{x}}_2 - \frac{n_1(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{n_1 + n_2} \right)^T \mathbf{A} \right)^T \left(\left(\mathbf{x}_i - \bar{\mathbf{x}}_2 - \frac{n_1(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{n_1 + n_2} \right)^T \mathbf{A} \right)$$

$$= \mathbf{A}^T \left(\frac{n_1 \Sigma_1 + n_2 \Sigma_2}{n_1 + n_2} + \frac{n_1 n_2^2 + n_2 n_1^2}{(n_1 + n_2)^3} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \right) \mathbf{A}$$

$$= \mathbf{A}^T \left(\frac{n_1 \Sigma_1 + n_2 \Sigma_2}{n_1 + n_2} + \frac{n_1 n_2}{(n_1 + n_2)^2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \right) \mathbf{A} = \mathbf{A}^T \Sigma_W \mathbf{A} + \boxed{\mathbf{A}^T \Sigma_B \mathbf{A}}$$

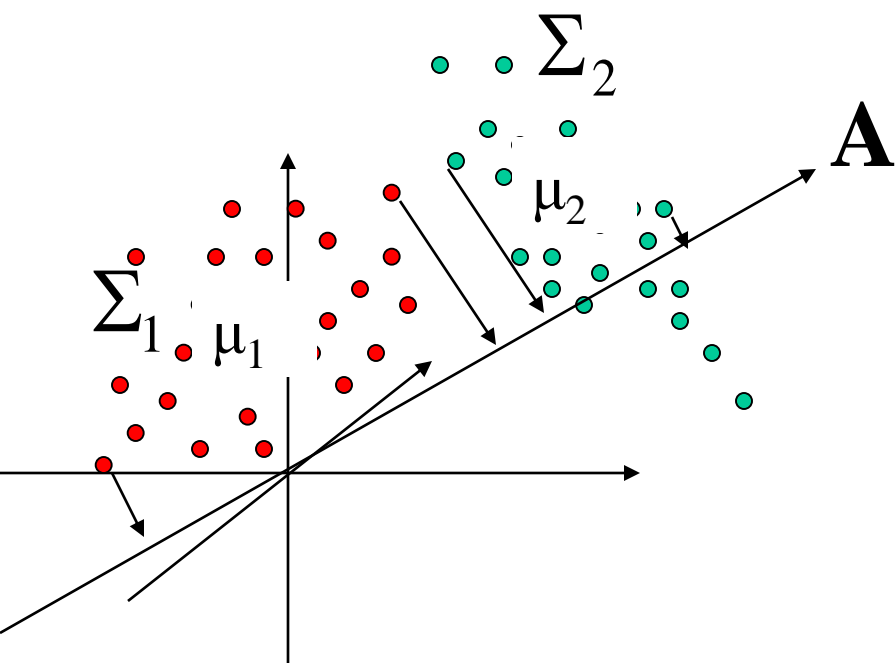
注意点

- ランクについて
 Σ_B はランク1
 Σ_W は, . . .

Σ_B^{-1} の計算は不可能.

識別に有利な特徴への変換

クラス内分散を小さくし、クラス間分散を大きくする一次元特徴を求める



$$\Sigma_W = \frac{n_1 \Sigma_1 + n_2 \Sigma_2}{n_1 + n_2}$$

$$\Sigma_B = \frac{n_1 n_2 (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T}{(n_1 + n_2)^2}$$

クラス内分散 $\mathbf{A}^T \Sigma_W \mathbf{A}$

クラス間分散 $\mathbf{A}^T \Sigma_B \mathbf{A}$

フィッシャー比 $\frac{\mathbf{A}^T \Sigma_B \mathbf{A}}{\mathbf{A}^T \Sigma_W \mathbf{A}}$ の最大化

識別に有利な特徴への変換： 判別分析（定式化）

A の大きさが変わっても $\frac{A^T \Sigma_B A}{A^T \Sigma_W A}$ の値は変わらないので、

$A^T \Sigma_W A = 1$ という条件の下での $A^T \Sigma_B A$ の最大化問題と見なす。

Lagrange の未定係数法により、次の目的関数が得られる。

$$J(A) = A^T \Sigma_B A - \lambda (A^T \Sigma_W A - 1)$$

この目的関数の最大化により目的の軸 A が決定できる。

識別に有利な特徴への変換： 判別分析（解き方）

$$\frac{\partial}{\partial A} J(A) = 2\Sigma_B A - 2\lambda\Sigma_W A = 0 \quad \text{なので、}$$

$$\Sigma_B A = \lambda\Sigma_W A \quad \text{が得られ、これを整理して}$$

$$\Sigma_W^{-1}\Sigma_B A = \lambda A \quad \text{となる。}$$

$\Sigma_W^{-1}\Sigma_B$ の最大固有値が J の最大値となり、

これに対応する固有ベクトルが、 A になる。

線形判別法： 多クラスの場合

多クラスの場合、Aは複数のベクトルとなる。

$$\Sigma_W = \sum_{i=1}^c P(\omega_i) \Sigma_i \quad \Sigma_B = \sum_{i=1}^c \sum_{j<i} P(\omega_i) P(\omega_j) (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j) (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)^T$$

これらを最大化する問題は、同じ問題に帰着する。

$\Sigma_W^{-1} \Sigma_B$ の固有ベクトルが、Aになる。

パターン認識 —統計的識別—

担当: 和田 俊和
部屋 A513

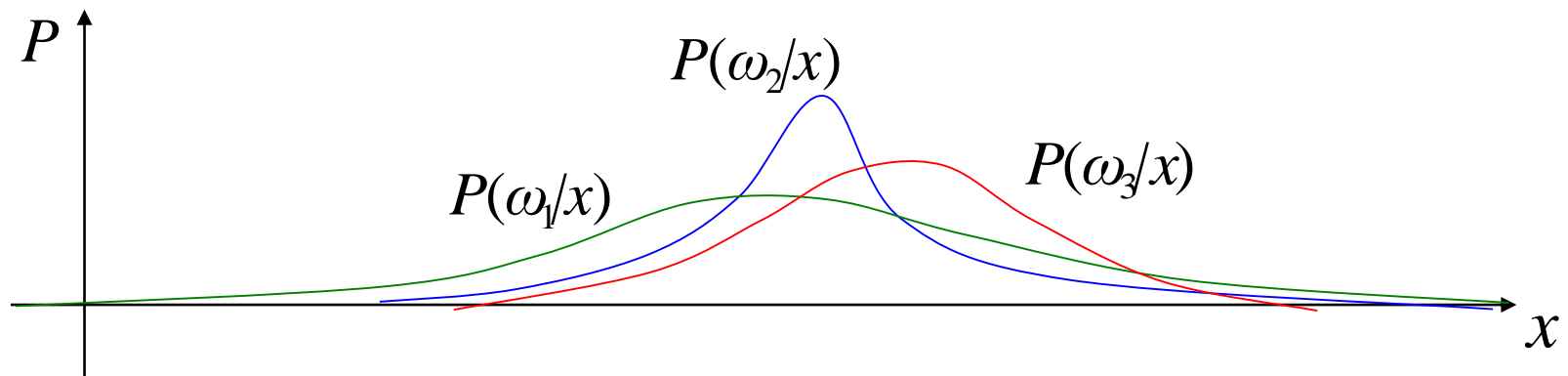
Email twada@ieee.org

講義資料は<http://vrl.sys.wakayama-u.ac.jp/PRA/WPR-2>

判別分析と統計的識別

統計的識別とは

- データ x が与えられたときに、それがああるクラス ω_i に属する確率 $P(\omega_i/x)$ を求める。
- 各クラスについて、 $P(\omega_i/x)$ を計算し、それを元にしてユーザの定義した**損失関数**を最小にする ω_i に x を分類する。



統計的パターン認識のための 数学的準備

クラス: $\omega_i (i = 1, \dots, c)$

入力データ: x

事前確率: $P(\omega_i), P(x)$

事後確率: $P(\omega_i/x), P(x|\omega_i)$

$$\sum_{i=1}^c P(\omega_i) = 1, \quad \sum_{i=1}^c P(\omega_i/x) = 1$$

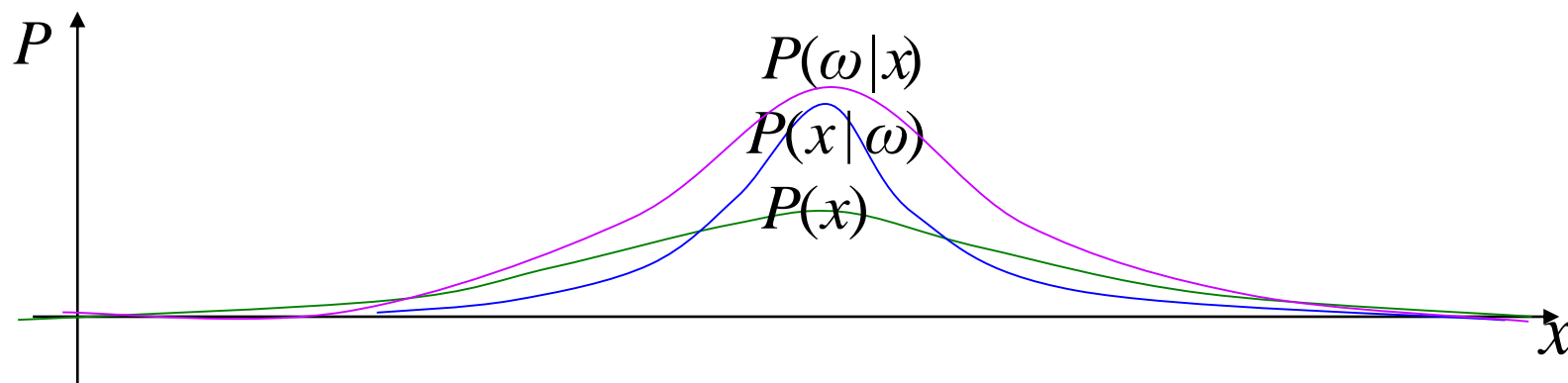
$$P(x) = \sum_{i=1}^c P(\omega_i)P(x|\omega_i)$$

Bayes 則 (定理) : $P(\omega_i | x)$ の計算

パターン x が入力されたとき、それがクラス ω_i に属する確率

クラス ω_i に属するパターン x が生起する確率

$$P(\omega_i | x) = \frac{P(x | \omega_i) P(\omega_i)}{P(x)}$$



識別基準:

Bayes決定則 = 期待損失の最小化

期待損失: (ω_j に属するパターン x を分類する際の損失)

$$L(\omega_j | x) = \sum_{i=1}^c l(\omega_j | \omega_i) P(\omega_i | x)$$

0-1 損失基準:
$$l(\omega_j | \omega_i) = \begin{cases} 0 & \text{if } j = i \\ 1 & \text{otherwise} \end{cases}$$

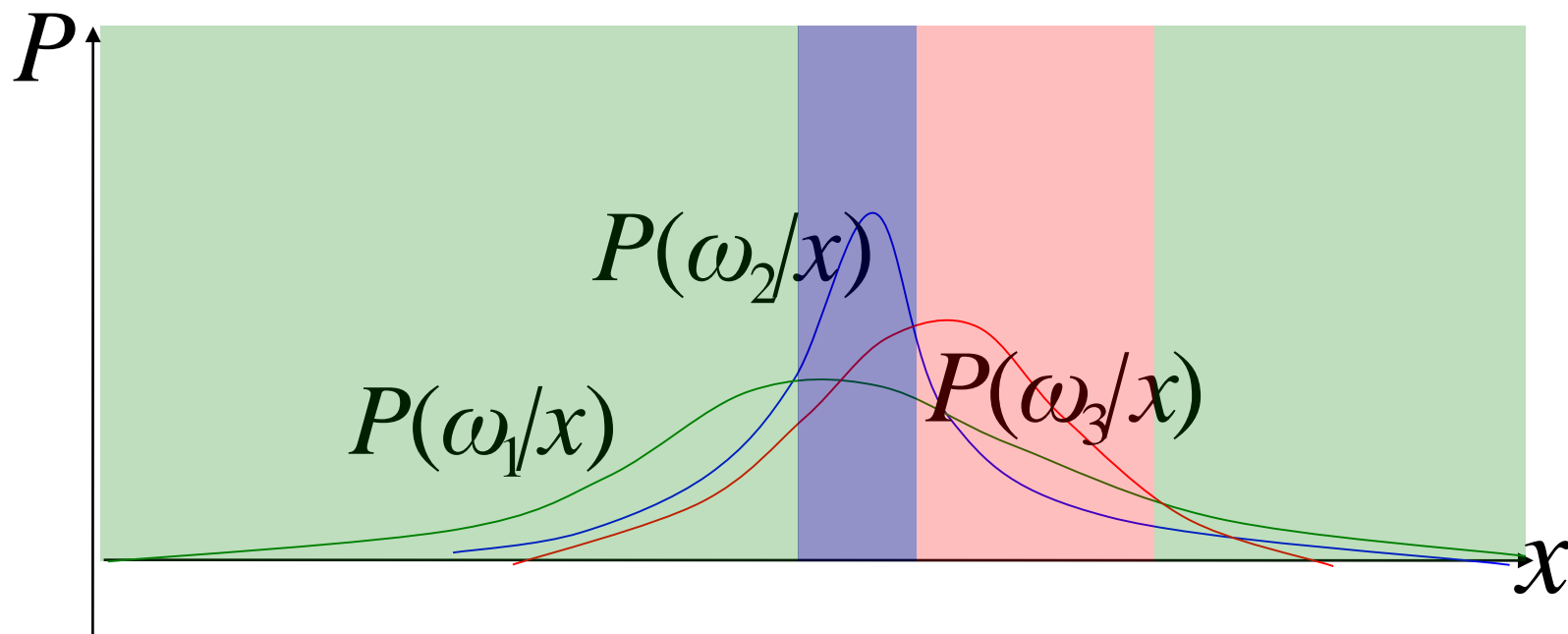
期待損失:

$$L(\omega_j | x) = \sum_{i \neq j}^c P(\omega_i | x) = 1 - P(\omega_j | x)$$

$L(\omega_j | x)$ を最小化することは $P(\omega_j | x)$ を最大化すればよい。
これは、事後確率を最大化するように識別を行えば良いことを意味している。

0-1 損失基準に基づく識別

- $P(\omega_j/x)$ の最大値を与える ω_i に x を分類する。



他の損失基準

期待損失

$$L(\omega_j | x) = \sum_{i=1}^c l(\omega_j | \omega_i) P(\omega_i | x)$$

識別器の出力: $\Psi(x) = (y_1, \dots, y_c)$

教師信号: $\mathbf{t} = (0, \dots, 1, \dots, 0)$

平均2乗誤差基準:

$$L(\Psi | x) = \sum_{i=1}^c P(\omega_i | x) \int \|\Psi(x; \theta) - \mathbf{t}\|^2 p(x | \omega_i) dx$$

連続損失基準

期待損失

$$L(\omega_j | x) = \sum_{i=1}^c l(\omega_j | \omega_i) P(\omega_i | x)$$

識別器の出力:

$$\Psi(x; \theta) = (g_1(x; \theta), \dots, g_c(x; \theta))$$

誤分類尺度

$$d_i(x) = \frac{1}{m_i} \sum_{j, g_j > g_i} g_j(x; \theta) - g_i(x; \theta)$$

[甘利]:

誤分類尺度

$$d_i(x) = -g_i(x; \theta) + \left[\frac{1}{c-1} \sum_{j \neq i} g_j(x; \theta)^\eta \right]^{1/\eta}$$

[Juang, Katagiri]:

Bayes 決定則を適用する上での問題
確率 $P(x|\omega_i)$ をどうやって求めるか

$$P(\omega_i | x) = \frac{P(x | \omega_i) P(\omega_i)}{P(x)}$$

$P(x|\omega_i)$ を確率密度関数 $p(x|\omega_i)$ で代替する。

$$P(\omega_i | x) = \frac{p(x | \omega_i) P(\omega_i)}{\sum_{j=1}^c P(\omega_j) p(x | \omega_j)}$$

Bayes 決定則に基づく識別

$$P(\omega_i | x) = \frac{p(x | \omega_i) P(\omega_i)}{\sum_{j=1}^c P(\omega_j) p(x | \omega_j)}$$

ある x に対して決まる分母は同じ値になるので、

$$g_i(x) = p(x | \omega_i) P(\omega_i)$$

の大小関係を調べて、識別を行えば良い。

確率密度関数の例

d次元正規分布

$$p(x | \omega_i) = \frac{\exp\left\{-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right\}}{(2\pi)^{d/2} |\Sigma_i|^{1/2}}$$

$$\int_{x \in \chi_i} p(x | \omega_i) dx = 1$$

例: 1次元正規分布

$$p(x | \omega_i) = \frac{\exp\left\{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right\}}{\sqrt{2\pi}\sigma}$$

Bayes の定理に基づく識別

仮定1 各クラスに属するパターンの事前確率 $P(\omega_i)$ が同じであるとすれば、

$$g_i(x) = p(x | \omega_i)$$

の大小関係を調べて、識別を行えば良いことになる。
正規分布の場合、その対数を求めると、

$$\log P(x | \omega_i) = -\frac{1}{2} \left(d \log 2\pi + \log(|\Sigma_i|) + (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right)$$

が得られる。このうち、クラスに依存しない項を除けば、

$$\log(|\Sigma_i|) + (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \quad \text{となる。}$$

Bayes の定理に基づく識別 と線形識別関数

仮定2 さらに、各クラスの共分散行列の行列式の値が等しいとすれば、

$$D_M^2(x, \mu_i) = (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)$$

の値が小さければ、 x と μ_i が似ていることになる。
これを「マハラノビス距離」という。

仮定3 各クラスの共分散行列自体も等しいとすれば、

$$g'_i(x) = 2\mu_i^T \Sigma^{-1} x - \mu_i^T \Sigma^{-1} \mu_i$$

となり、線形識別関数が得られる。(仮定1が無くても仮定3のみで、下記の線形識別関数が得られる。)

$$g_i(x) = 2\mu_i^T \Sigma^{-1} x - \mu_i^T \Sigma^{-1} \mu_i + 2 \log P(\omega_i)$$

確率密度関数の母数の推定(1)

確率密度関数 $p(x; \theta)$ の形は分かっているが、分布形状を決定するパラメータ(母数) θ が未知である場合、学習サンプルから θ を決定したい。

$$p(\chi, \theta) = \prod_{k=1}^n p(x_k; \theta_i)$$

θ の関数とみなしたとき、これを「尤度関数」と呼ぶ

これが、最適な θ で、停留すると仮定し、

$$\frac{\partial}{\partial \theta} p(\chi, \theta) = 0$$

を満たす θ を求める。(最尤推定)

これは、全ての x_k が同時に生起するという確率(同時生起確率)であり、「全ての学習サンプルを説明する上で、最も尤もらしい分布の母数を推定する」ことを意味している。

パラメータの最尤推定 正規分布の場合(Σ, μ)

$p(\chi, \theta)$ 最大化の結果

共分散行列:
$$\Sigma_i = \frac{1}{n_i} \sum_{x \in \chi_i} (x - \mu_i)(x - \mu_i)^T$$

平均:
$$\mu_i = \frac{1}{n_i} \sum_{x \in \chi_i} x$$

確率密度関数の母数の推定(2)

不偏推定

すべての可能な χ に対する θ の推定量の期待値が真の値に対して偏らないように推定する。(不偏推定)

$$E_{\chi}(\theta(\chi)) - \theta_{true} = 0$$

正規分布の場合

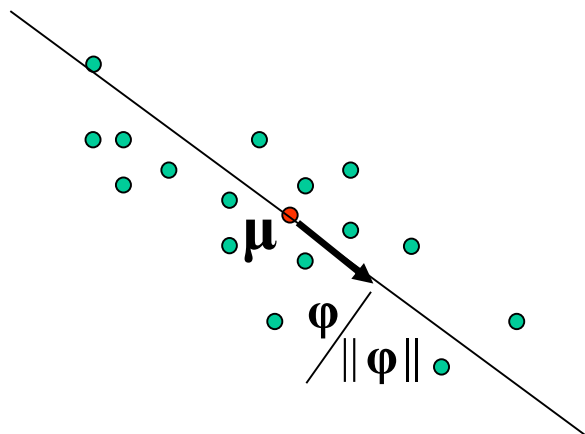
共分散行列: $\Sigma_i = \frac{1}{n_i - 1} \sum_{x \in \chi_i} (x - \mu_i)(x - \mu_i)^T$

平均: $\mu_i = \frac{1}{n_i} \sum_{x \in \chi_i} x$

正規分布の推定例

主成分分析

- 共分散行列 Σ の性質



$$S = \begin{bmatrix} \hat{e}_1 & S_1^2 & S_{12} & \cdots & S_{1n} \\ \hat{e}_2 & S_2^2 & S_{22} & \cdots & S_{2n} \\ \hat{e}_3 & S_3^2 & S_{32} & \cdots & S_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{e}_n & S_n^2 & S_{n2} & \cdots & S_n^2 \end{bmatrix}$$

対称行列なので次式による
対角化が可能

$$\Sigma = V\Lambda V^T$$

$$V = [\varphi_1 \quad \varphi_2 \quad \cdots \quad \varphi_N]$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

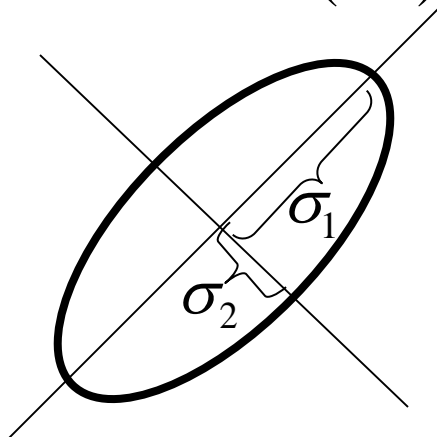
共分散行列の性質

$$D_M^2(x, \mu_i) = (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)$$

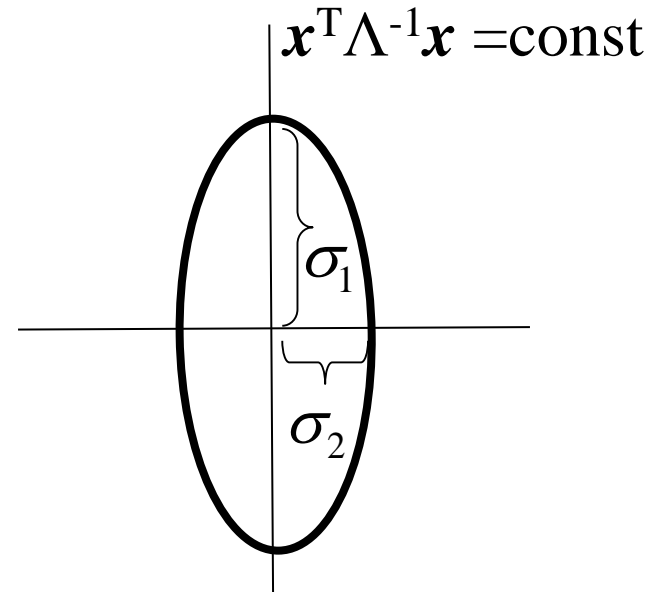
マハラノビス距離が一定の曲面(曲線)は正規分布の確率密度が一定の面(線)であり、その形状は楕円面(楕円)になる

$S^{-1} = VL^{-1}V^T$ の意味

$$\mathbf{x}^T \Sigma^{-1} \mathbf{x} = (\mathbf{V}\mathbf{x})^T \Lambda^{-1} (\mathbf{V}\mathbf{x}) = \text{const}$$



空間をVによって回転させるとこのように見える。



主成分分析

- 共分散行列 Σ の性質

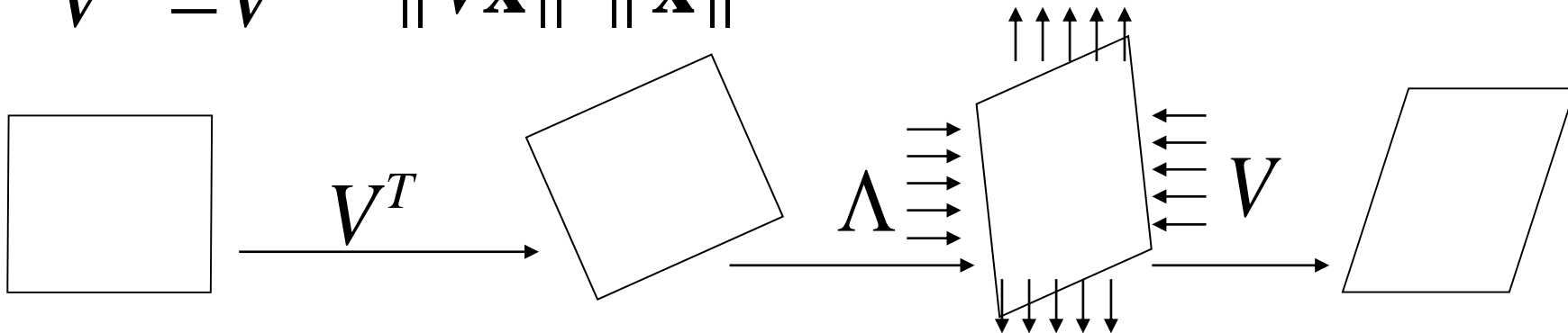
$$\Sigma = V\Lambda V^T$$

$$V = [\boldsymbol{\varphi}_1 \quad \boldsymbol{\varphi}_2 \quad \cdots \quad \boldsymbol{\varphi}_N]$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

V は直交行列(回転を表している):

$$V^T = V^{-1} \quad \|V\mathbf{x}\| = \|\mathbf{x}\|$$



正規分布の意味

- d次元正規分布

共分散行列を使った二次形式

$$p(x | \omega_i) = \frac{\exp\left\{-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)\right\}}{(2\pi)^{d/2} |\Sigma_i|^{1/2}}$$

↑
共分散行列の行列式
(正規化項: 分布の広がりを表している)

主成分分析によって何が分かるか

- 分散の大きくなる軸の向き ϕ_i
- その軸方向の偏差 $\sqrt{\lambda_i}$ 分散 λ_i
- 共分散行列のランクが r である場合、次式が成り立つ

$$\mathbf{x} = \sum_{i=1}^r (\phi_i \cdot (\mathbf{x} - \bar{\mathbf{x}})) \phi_i + \bar{\mathbf{x}}$$

- r 未満の数 q に対する直交展開

$$\mathbf{x}' = \sum_{i=1}^q (\phi_i \cdot (\mathbf{x} - \bar{\mathbf{x}})) \phi_i + \bar{\mathbf{x}}$$

を計算する際に誤差 $\|\mathbf{x} - \mathbf{x}'\|$ を最小化する基底が求まっている。