

はじめてのパターン認識

4章 確率モデルと識別関数

茨城大学理工学研究科

國井慎也

はじめに

- データの特徴間には相関がある
 - 100円硬貨の重さと直径
 - 使い古されて周りがすり減った硬貨 → 重さと直径がともに小さくなる
- $C = \{\text{新しい、古い}\}$ は、どちらか一方の特徴量があればよい



より識別精度を高めるためには、複数の特徴量を使う必要がある

特徴間には相関がない方が望ましい

- 平均ベクトルと共分散行列
 - 確率分布のパラメータ
 - 特徴量の線形変換により相関をなくせる

平均ベクトルと共分散行列

- 観測データ

- $\boldsymbol{x} = (x_1, x_2, \dots, x_d)^T$

- 平均ベクトル

- $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_d)^T = (E\{x_1\}, E\{x_2\}, \dots, E\{x_d\})^T$

- 共分散行列

- $\boldsymbol{\Sigma} = \text{Var}\{\boldsymbol{x}\} = E\{(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T\}$

$$\begin{aligned} &= \begin{pmatrix} E\{(x_1 - \mu_1)(x_1 - \mu_1)^T\} & \cdots & E\{(x_1 - \mu_1)(x_d - \mu_d)^T\} \\ \vdots & \ddots & \vdots \\ E\{(x_d - \mu_d)(x_1 - \mu_1)^T\} & \cdots & E\{(x_d - \mu_d)(x_d - \mu_d)^T\} \end{pmatrix} \\ &= (\sigma_{ij}) = \begin{cases} i = j \text{ 分散} \\ i \neq j \text{ 共分散} \end{cases} \end{aligned}$$

標準化

- 測定単位の取り方で特徴量の値が変化

- 測定単位の影響を除く必要がある

➡ 標準化：データの平均 0、分散 1

- 特徴 x の線形変換を $y = ax + b$

- $E\{y\} = E\{ax + b\} = aE\{x\} + b = a\mu + b$

- $Var\{y\} = E\{(y - E\{y\})^2\} = E\{(ax + b - a\mu - b)^2\}$
 $= a^2 Var\{x\} = a^2 \sigma^2$

- 線形変換によって標準化

- $z = \frac{x - \mu}{\sigma} \quad a = \frac{1}{\sigma} \quad b = -\frac{\mu}{\sigma}$

- $E\{z\} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0 \quad Var\{z\} = \left(\frac{1}{\sigma}\right)^2 \sigma^2 = 1$

観測データの無相関化

- 固有値問題

$$\Sigma s = \lambda s$$

- 得られた固有値 $\lambda_1, \lambda_2, \dots, \lambda_d$ に対応する固有ベクトル

$$S = (s_1, s_2, \dots, s_d)$$

- 観測データ x を S^T で写像することを考える

- $y = S^T x$

- $E\{y\} = E\{S^T x\} = S^T \mu$

$$\begin{aligned} \text{Var}\{y\} &= E\{(y - E\{y\})(y - E\{y\})^T\} \\ &= S^{-1} E\{(x - \mu)(x - \mu)^T\} S = S^{-1} \Sigma S \end{aligned}$$

$$S^{-1} \Sigma S = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{pmatrix}$$

白色化

- 固有値に相当する分だけ特徴量の標準偏差に違い
 - 白色化後の座礁系

- $u = (u_1, u_2, \dots, u_d)^T$

- $u = \Lambda^{-1/2} S^T (x - \mu)$

$\Lambda^{-1/2}$: Λ の対角平方根
を取った行列の逆行列

- 平均

$$E\{u\} = \Lambda^{-1/2} S^T (E\{x\} - \mu) = \Lambda^{-1/2} S^T (\mu - \mu) = 0$$

- 共分散行列

$$\text{Var}\{u\} = E\{uu^T\} = E\left\{\Lambda^{-1/2} S^T (x - \mu)(x - \mu)^T S \Lambda^{-T/2}\right\}$$

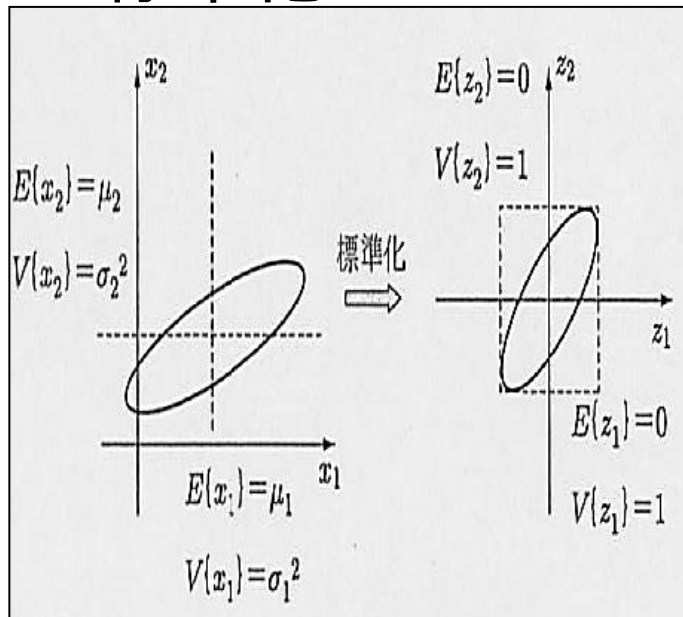
$$= \Lambda^{-1/2} S^{-1} E\left\{(x - \mu)(x - \mu)^T\right\} S \Lambda^{-T/2}$$

$$= \Lambda^{-1/2} S^{-1} \Sigma S \Lambda^{-T/2}$$

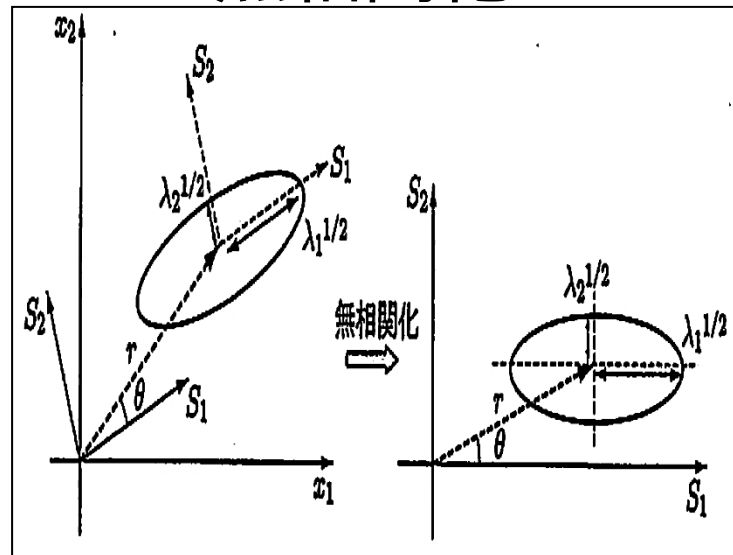
$$\text{Var}\{u\} = \Lambda^{-1/2} \Lambda \Lambda^{-T/2} = I$$

標準化、無相関化、白色化

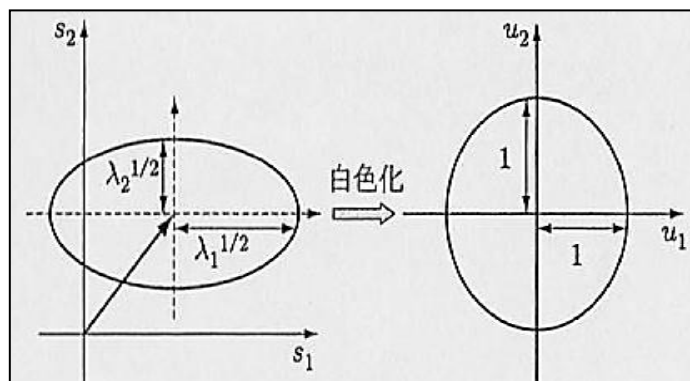
標準化



無相関化



白色化



確率モデル

- 学習データの分布を表現
 - パラメトリックモデル
 - 学習データから推定した統計量（平均、分散）を用いて構成した確率モデルを利用したモデル
 - ノンパラメトリックモデル
 - 学習データそのものを利用して表現するモデル

正規分布関数

一次元

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

多次元

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

正規分布の識別関数

- クラス*i*の条件付き確率

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

- ベイズ誤り率最小識別規則

– クラスの事前確率： $P(C_i)$

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

対数をとる

$$\ln P(C_i) - \frac{d}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i)$$

符号を反転する

$$g_i(\mathbf{x}) = (\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i) + \ln |\Sigma_i| - 2 \ln P(C_i)$$

$$\text{識別クラス} = \arg \min_i [g_i(\mathbf{x})]$$

正規分布の最尤推定

- 尤度
 - 標本データが、ある母集団から得られる確率のこと
 - 人為的に作成した複数のモデルの候補の中で、どのモデルが真のモデルに近いかを判断する必要がある。その際に尤度が用いられ、尤度が高いモデルを標本データに最も適していると判断する
 - 尤度を求めるには尤度関数が必要である

尤度関数

- 学習データ $x_i (i = 1.., N)$ パラメータ $\theta = (\theta_1, \dots, \theta_M)$ とする。このとき、真の確率モデル $f(x|\theta)$ に従う N 個の学習データの同時分布を考える。
- $L(\theta) = f(x_1, \dots, x_N|\theta) = \prod_{i=1}^N f(x_i|\theta)$
- この関数を尤度関数と呼ぶ

導出

- 微分を行う

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_i} = 0 \quad (i = 1, \dots, M)$$

- 一次元正規分布の平均値 μ と分散 σ^2 を求める

$$\begin{aligned} L(\mu, \sigma^2) &= f(x_1, \dots, x_N | \mu, \sigma^2) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\ &= (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2\right) \\ \mathcal{L}(\mu, \sigma^2) &= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \end{aligned}$$

導出 2

- 平均值

$$\frac{\partial \mathcal{L}(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \mu) = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

- 分散

$$\frac{\partial \mathcal{L}(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{1}{N} \frac{1}{\sigma^2} - \frac{2}{(2\sigma^2)^2} \sum_{i=1}^N (x_i - \mu)^2 = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2$$