



演習

分散・共分散行列

パターン認識論

高精度な認識へ向けて

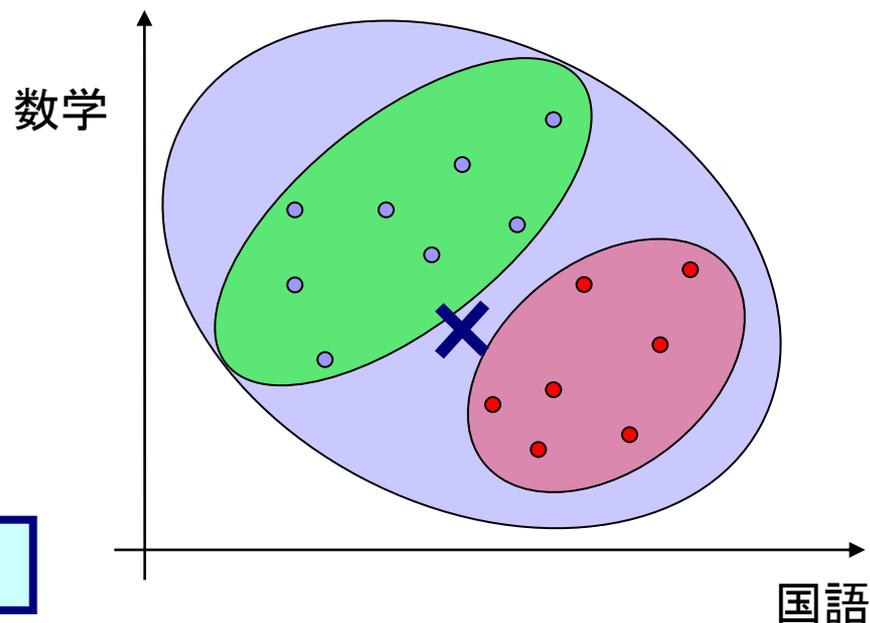
■ 前回まで

- ユークリッド距離
- 類似度
- 重み付きユークリッド距離

平均データの位置情報のみを考慮



分散情報の利用 ⇔ 認識精度の向上



分散・共分散行列

- 多次元データの分散の表現法のひとつ

$$S = \begin{pmatrix} S_{1,1} & S_{1,2} & S_{1,3} \\ S_{2,1} & S_{2,2} & S_{2,3} \\ S_{3,1} & S_{3,2} & S_{3,3} \end{pmatrix}$$

1,2,3次元それぞれの分散

異なる次元のデータ同士の共分散
(1-2次元、2-3次元など)

分散: ある次元のばらつき、共分散: 2つの次元間の関係

- データの対象によって3つの種類がある
 - 全共分散行列
 - クラス内共分散行列
 - クラス間共分散行列

各種共分散行列

- 全共分散行列: 全データにおける共分散行列

$$C_T = \frac{1}{N} \sum_{i=1}^N (x_i - m)(x_i - m)^T$$

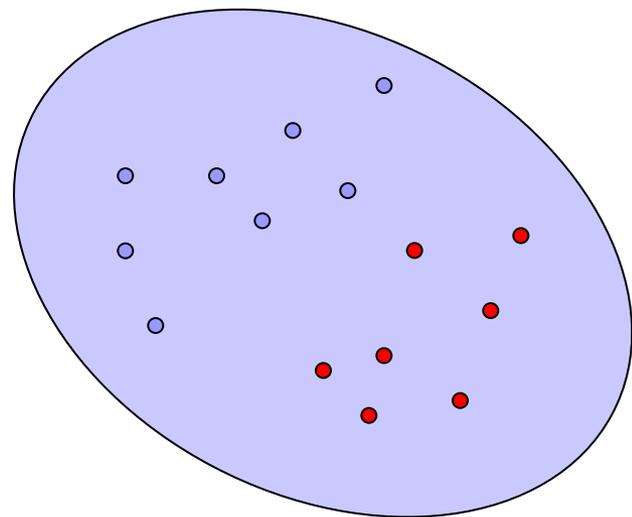
※ T は転置行列

入力データ集合

$$X = (x_1, x_2, \dots, x_N)$$

平均ベクトル

$$m = \frac{1}{N} \sum_{i=1}^N x_i$$



各種共分散行列

- クラス内共分散行列: 各クラスのデータにおける共分散行列

$$C_W = \frac{1}{N} \sum_{K=1}^C \sum_{i=1}^{N_K} (x_i - m_K)(x_i - m_K)^T$$

- クラス間共分散行列: クラス平均データ間の共分散行列

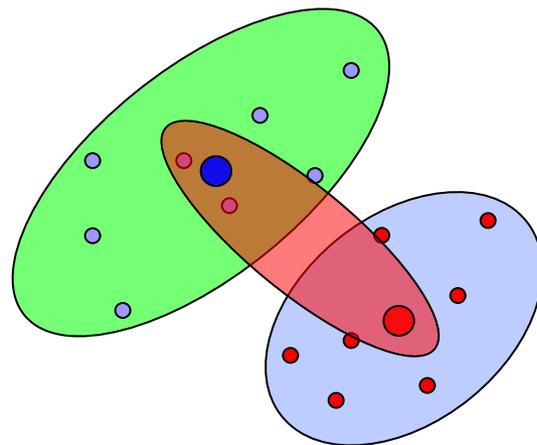
$$C_B = \frac{1}{N} \sum_{K=1}^C N_K (m_K - m)(m_K - m)^T$$

クラス番号

$$K = 1, 2, \dots, C$$

クラス平均ベクトル

$$m_K = \frac{1}{N_K} \sum_{i=1}^{N_K} x_i$$



共分散行列の計算例(全共分散)

$$\left. \begin{array}{l}
 x_1 = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \\
 x_2 = \begin{bmatrix} 8 \\ 6 \end{bmatrix} \\
 x_3 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \\
 x_4 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}
 \end{array} \right\} \begin{array}{l}
 \text{Class A} \\
 m_A = \begin{bmatrix} 7 \\ 6 \end{bmatrix} \\
 \text{Class B} \\
 m_B = \begin{bmatrix} 5 \\ 4 \end{bmatrix}
 \end{array} \right\} m = \begin{bmatrix} 6 \\ 5 \end{bmatrix}$$

全共分散行列

$$\frac{1}{4} \sum_{i=1}^4 (x_i - m)(x_i - m)^t$$

$$= \frac{1}{4} \left(\begin{array}{l}
 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix} \\
 + \begin{bmatrix} -2 \\ -1 \end{bmatrix} \begin{bmatrix} -2 & -1 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \begin{bmatrix} 0 & -1 \end{bmatrix}
 \end{array} \right)$$

$$= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

課題2

- データセットAにおいて、各共分散行列を出力するプログラムを作成し、実行せよ

- サンプルプログラムダウンロード

<http://www.it.ecei.tohoku.ac.jp/pattern/sample2.c>

- コンパイル、実行

- `gcc sample2.c -llapack -lblas -lf2c -lm`
- `./a.out 1 2D-class1_A.dat 2D-class2_A.dat 2D-test_A.dat`

- 正しい結果はこちら

全共分散行列		クラス内共分散		クラス間共分散	
486.30	94.01	206.70	96.41	279.59	-2.39
94.01	110.75	96.41	110.73	-2.39	0.02

※小数第3位以下切り捨て