知能機械と自然言語処理
知能機械部　第6回

ソフトウェア情報学部

Goutam Chakraborty

d次元の教師データ

1番目の特色　2番目の特色　特色

教師データ

$$
\begin{array}{lc}
\text{1番目のデータ} & \vec{x^1} \\
\text{2番目のデータ} & \vec{x^2} \\
 & \vdots \\
\text{i番目のデータ} & \vec{x^i} \\
 & \vdots \\
\text{n番目のデータ} & \vec{x^n}
\end{array}
\quad
\begin{bmatrix}
< x_1^1 & x_2^1 & \cdots & x_k^1 & \cdots & x_d^1 > \\
< x_1^2 & x_2^2 & \cdots & x_k^2 & \cdots & x_d^2 > \\
\vdots & \vdots & \ddots & \vdots & & \vdots \\
< x_1^i & x_2^i & \cdots & x_k^i & \cdots & x_d^i > \\
\vdots & \vdots & & \vdots & \ddots & \vdots \\
< x_1^n & x_2^n & \cdots & x_k^n & \cdots & x_d^n >
\end{bmatrix}
\quad
\begin{bmatrix}
0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 1
\end{bmatrix}
$$

特色ベクトルの中の
成分の順番が重要！

d次元の正規分布の確率密度関数

一次元の正規分布の確率密度関数

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

d次元の場合

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\left(\vec{x}-\vec{\mu}\right)^T \Sigma^{-1}\left(\vec{x}-\vec{\mu}\right)\right)$$

d : 次元数(dimension)

Σ : 相関行列

相関行列の求め方

$$
\begin{array}{ccccc}
\vec{x^1} & \vec{x^2} & \cdots & \vec{x^{3000}} & \\
\begin{bmatrix} 46.1 \\ 180 \end{bmatrix} & \begin{bmatrix} 31.6 \\ 241 \end{bmatrix} & \cdots & \begin{bmatrix} x_1^{3000} \\ x_2^{3000} \end{bmatrix} & \begin{array}{l} \text{1番目の特色の平均} \mu_1 = 45 \\ \text{2番目の特色の平均} \mu_2 = 200 \end{array}
\end{array}
$$

$$x_d^n - \mu_d = \hat{x}_d^n$$

平均を引くことで変化後の平均は0になる

相関行列(2)

特色1の分散

$$
n\Sigma = \begin{bmatrix} \hat{x}_1^1 & \hat{x}_1^2 & \cdots & \hat{x}_1^n \\ \hat{x}_2^1 & \hat{x}_2^2 & \cdots & \hat{x}_2^n \end{bmatrix}
\begin{bmatrix} \hat{x}_1^1 & \hat{x}_2^1 \\ \hat{x}_1^2 & \hat{x}_2^2 \\ \vdots & \vdots \\ \hat{x}_1^n & \hat{x}_2^n \end{bmatrix}
= \begin{bmatrix} \sum_{i=1}^n (\hat{x}_1^i)^2 & \sum_{i=1}^n \hat{x}_1^i \hat{x}_2^i \\ \sum_{i=1}^n \hat{x}_2^i \hat{x}_1^i & \sum_{i=1}^n (\hat{x}_2^i)^2 \end{bmatrix}
$$

特色2の分散

$$
\begin{bmatrix} \hat{x}_1^1 & \hat{x}_1^2 & \cdots & \hat{x}_1^n \\ \hat{x}_2^1 & \hat{x}_2^2 & \cdots & \hat{x}_2^n \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}_d^1 & \hat{x}_d^2 & \cdots & \hat{x}_d^n \end{bmatrix}
\begin{bmatrix} \hat{x}_1^1 & \hat{x}_2^1 & \cdots & \hat{x}_d^1 \\ \hat{x}_1^2 & \hat{x}_2^2 & \cdots & \hat{x}_d^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}_1^n & \hat{x}_2^n & \cdots & \hat{x}_d^n \end{bmatrix}
= \begin{bmatrix} \sum_{i=1}^n (\hat{x}_1^i)^2 & \sum_{i=1}^n \hat{x}_1^i \hat{x}_2^i & \cdots & \sum_{i=1}^n \hat{x}_1^i \hat{x}_d^i \\ \sum_{i=1}^n \hat{x}_2^i \hat{x}_1^i & \sum_{i=1}^n (\hat{x}_2^i)^2 & \cdots & \sum_{i=1}^n \hat{x}_2^i \hat{x}_d^i \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n \hat{x}_d^i \hat{x}_1^i & \sum_{i=1}^n \hat{x}_d^i \hat{x}_2^i & \cdots & \sum_{i=1}^n (\hat{x}_d^i)^2 \end{bmatrix}
$$

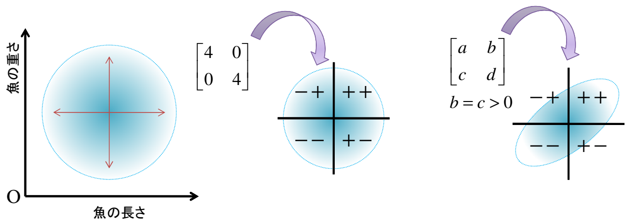
課題

- Nishin2D-300.txt, Sanma2D-500.txt から相関行列 Σ を求める。データは下記のURLに保存されている

http://www.chishiki.soft.iwate-pu.ac.jp/ISNLP/isnlp.html

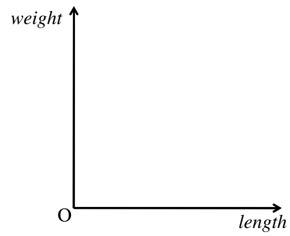
相関行列とデータの分布の関連

- 相関行列から
  - ➔固有ベクトル，固有値を求めると特色空間内のデータ分布の形が分かる
    - ➔主成分分析（**P**rincipal **C**omponent **A**nalysis）

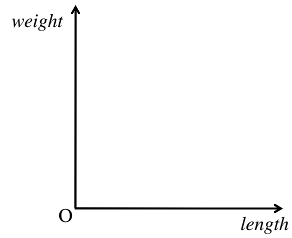


相関行列とデータの形の練習

- 問題１
  - $\begin{bmatrix}4 & 0\\0 & 1\end{bmatrix}$を以下に図示しろ



- 問題２
  - ▶ $\begin{bmatrix}4 & 2\\2 & 1\end{bmatrix}$を以下に図示しろ



相関行列-データの次元数(特色の数)が増えると分布の形を考え難い

- 自由課題

$\begin{bmatrix}5 & -2 & 1\\-2 & 2 & 5\\1 & 5 & 1\end{bmatrix}$について、どのようになるか考えよう！

行列式の求め方

行列 $\sum=\begin{bmatrix}4 & 3\\3 & 1\end{bmatrix}$ の行列式は

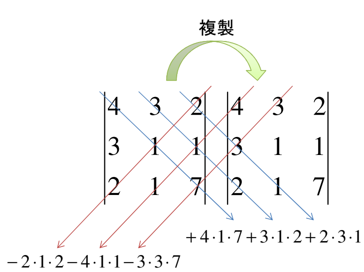
$|\Sigma|$ について、求めるには？

$$|\Sigma|=\det(\Sigma)=\begin{vmatrix}4 & 3\\3 & 1\end{vmatrix}=|4\cdot 1-3\cdot 3|=5$$

問題：以下の行列式を求めよう！

$$\det(\Sigma)=\begin{vmatrix}4 & 3 & 2\\3 & 1 & 1\\2 & 1 & 7\end{vmatrix}=\ ?$$

行列式（2）



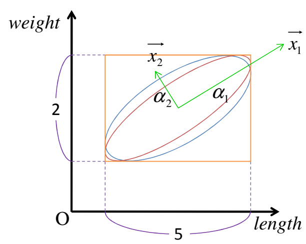
$$\therefore\ \det(\Sigma)=\begin{vmatrix}4 & 3 & 2\\3 & 1 & 1\\2 & 1 & 7\end{vmatrix}=31$$

固有値，固有ベクトル（1）

$\boxed{\begin{bmatrix}5 & 1\\1 & 2\end{bmatrix}}$ $\boxed{\begin{bmatrix}5 & 2\\2 & 2\end{bmatrix}}$

固有値 α_1, α_2

固有ベクトル $\vec{x}_1, \vec{x}_2$



$\begin{bmatrix}X_{11} & 高\\高 & X_{22}\end{bmatrix}$ $\alpha_1 >> \alpha_2$

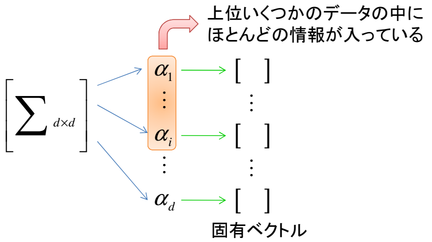
相関**高**⇒長さと重さの関係がはっきり分かる

- 両方の特色が不要
- 片方の特色で他の特色の値が分かる
  - ➔データの圧縮

$\begin{bmatrix}● & 高\\高 & X_{22}\end{bmatrix}$

一つのデータだけでOK！

固有値，固有ベクトル（2）



- 自由課題
  - にしんとさんまのデータについて、固有値・固有ベクトルを求めよ。