パターン認識
Pattern Recognition

佐藤真一
Shin'ichi Satoh

国立情報学研究所
National Institute of Informatics

Apr 18, 2023

本講義について

- 講義のページ: https://research.nii.ac.jp/~satoh/utpr/
- 単位は最終レポートと宿題によって評価予定 (最終レポートの提出は必須、宿題は全 7 回のうち 3 回の提出必須)
- 出席は取らない

スケジュール (予定)

- 4/11 講義概要、ベイズ決定測、確率分布
- 4/18 確率変数、確率変数ベクトル、正規分布
- 4/25 確率分布パラメータ推定、識別関数
- 5/2 ノンパラメトリック分布推定、パルゼン窓、k 近傍推定
- 5/9 k 近傍識別器、識別誤り推定
- 5/16 ベイズ誤り推定、識別誤り推定、交差判定法、ブートストラップ法
- 5/23 線形識別器、パーセプトロン、最小自乗誤差 (MSE) 識別器、Widrow-Hoff 則
- 5/30 ニューラルネットワーク、深層学習
- 6/6 サポートベクターマシン
- 6/13 直交展開、固有値展開
- 6/20 休講予定
- 6/27 クラスタリング、デンドログラム、凝集型 (aggromerative) クラスタリング、k-means
- 7/4 グラフ、ノーマライズドカット、スペクトラルクラスタリング、ラプラシアンアイゲンマップ
- 7/11 予備日

本日の予定

- ベイズ決定則により確率分布さえわかれば識別ができることが分かった
- 本日は正規分布などパラメータで記述できる確率分布のパラメータ推定法について検討する
- 特に標本からパラメータを推定する方法、推定したパラメータの統計的特性について考える
- 空間の線形変換についても考える
- 統計的に主たる特徴量の選択や、特徴量間の相関を減ずるなどの目的のため使われる

確率変数ベクトルと分布

パターン認識システムの入力として確率変数ベクトル を考える:

$$\mathbf{X} = [\mathbf{x}_1\, \mathbf{x}_2\, \cdots\, \mathbf{x}_n]^T.$$

確率分布関数:

$$P(x_1, \ldots, x_n) = Pr\{\mathbf{x}_1 \le x_1\ \ldots, \mathbf{x}_n \le x_n\}$$

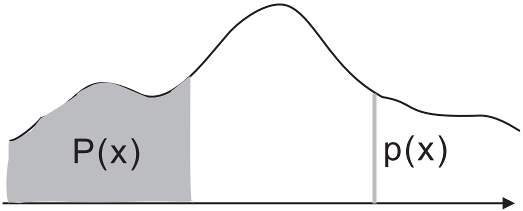
$$P(X) = Pr\{\mathbf{X} \le X\}$$

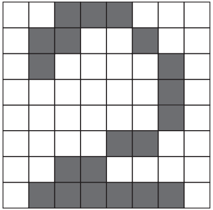
確率密度関数:

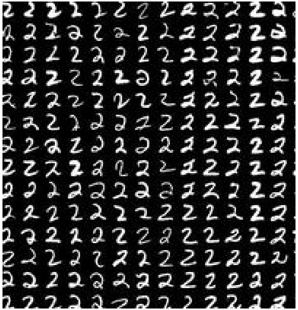
$$p(X) = \lim_{\substack{\Delta x_1 \to 0 \\ \vdots \\ \Delta x_n \to 0}} \frac{Pr\{x_1 < \mathbf{x}_1 \le x_1 + \Delta x_1, \ldots, x_n < \mathbf{x}_n \le x_n + \Delta x_1\}}{\Delta x_1 \cdots \Delta x_n}$$

$$= \frac{\partial^n P(X)}{\partial x_1 \ldots \partial x_n}$$

確率変数ベクトルと分布







確率変数ベクトルと分布

$$p(X) = \frac{\partial^n P(X)}{\partial x_1 \dots \partial x_n}$$
$$P(X) = \int_{-\infty}^{X} p(Y) dY = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} p(y_1, \dots, y_n) dy_1 \dots dy_n$$

例 1 コインを投げた時、**x** は表 (H) か裏 (T) $P(H) = P(T) = \frac{1}{2}$.

例 2 **x** は一様分布

$$p(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases}$$

例 3 **x** は正規分布

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

条件付確率密度、事前確率、事後確率

条件付確率密度

$$p(X|\omega_i) = p_i(X)$$

事前確率

$$P(\omega_i) = P_i$$

(条件なし) 確率密度

$$p(X) = \sum_i p_i(X)P_i$$

事後確率

$$P(\omega_i|X) = \frac{p_i(X)P_i}{p(X)}$$

分布のパラメータ

確率変数ベクトル **X** は、その確率密度関数により記述できる。しかし、関数を完全に表現するのは困難。その代わりに、パラメータにより記述することを考える。

期待値

$$M = E\{\mathbf{X}\} = \int Xp(X)dX$$

$$m_i = \int x_i p(X)dX = \int_{-\infty}^{\infty} x_i p(x_i)dx_i$$

$$p(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(X)dx_1 \cdots dx_{i-1}dx_{i+1} \cdots dx_n$$

(周辺密度)

条件付期待値

$$M_i = E\{\mathbf{X}|\omega_i\} = \int Xp_i(X)dX$$

共分散行列:

$$\begin{aligned}
\Sigma &= E\{(\mathbf{X} - M)(\mathbf{X} - M)^T\} \\
&= E\left\{\begin{bmatrix} \mathbf{x}_1 - m_1 \\ \vdots \\ \mathbf{x}_n - m_n \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - m_1 & \dots & \mathbf{x}_n - m_n \end{bmatrix}\right\} \\
&= E\left\{\begin{bmatrix} (\mathbf{x}_1 - m_1)(\mathbf{x}_1 - m_1) & \dots & (\mathbf{x}_1 - m_1)(\mathbf{x}_n - m_n) \\ \vdots & & \vdots \\ (\mathbf{x}_n - m_n)(\mathbf{x}_1 - m_1) & \dots & (\mathbf{x}_n - m_n)(\mathbf{x}_n - m_n) \end{bmatrix}\right\}
\end{aligned}$$

分布のパラメータ

共分散行列:

$$\Sigma = \begin{bmatrix} E\{(\mathbf{x}_1 - m_1)(\mathbf{x}_1 - m_1)\} & \dots & E\{(\mathbf{x}_1 - m_1)(\mathbf{x}_n - m_n)\} \\ \vdots & & \vdots \\ E\{(\mathbf{x}_n - m_n)(\mathbf{x}_1 - m_1)\} & \dots & E\{(\mathbf{x}_n - m_n)(\mathbf{x}_n - m_n)\} \end{bmatrix}$$

$$= \begin{bmatrix} c_{11} & \dots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \dots & c_{nn} \end{bmatrix}$$

特徴:

- 共分散行列は分散と共分散で構成される
- 共分散行列は対称行列

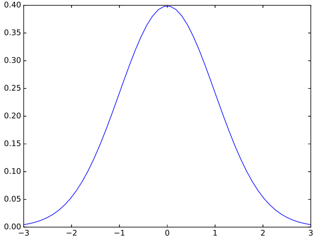
正規分布

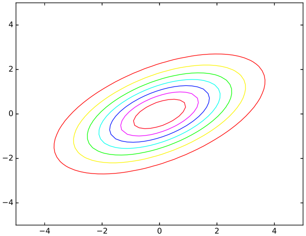
$$N_x(M, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}(X - M)^T\Sigma^{-1}(X - M))$$

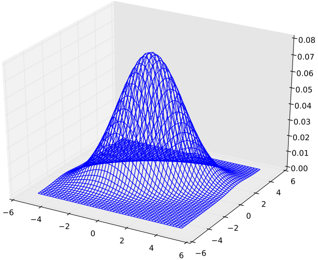
特徴:

- M と Σ は正規分布の 十分統計量
- $\mathbf{x}_i$ が無相関ならば独立でもある
  - $\mathbf{x} \sim N_x(0, \sigma)$ と $\mathbf{y} = \mathbf{x}^2$ とは無相関だが独立ではない (証明せよ).
  - $\mathbf{X}$, $\mathbf{Y}$ が無相関 $\rightleftharpoons E\{\mathbf{XY}\} = E\{\mathbf{X}\}E\{\mathbf{Y}\}$
  - $\mathbf{X}$, $\mathbf{Y}$ が独立 $\rightleftharpoons P(\mathbf{X} = x, \mathbf{Y} = y) = P(\mathbf{X} = x)P(\mathbf{Y} = y)$
- 正規分布の周辺密度も正規分布
- 中心極限定理による物理的正当化
  - 独立同一分布 (i.i.d.) に従う確率過程の和は正規分布に収束する

正規分布の例 (normdist.py)







マハラノビス距離

$$
\begin{aligned}
d^2(X) &= (X-M)^T\Sigma^{-1}(X-M) \\
&= tr\{\Sigma^{-1}(X-M)(X-M)^T\} \\
&= \sum_i \sum_j h_{ij}(x_i - m_i)(x_j - m_j)
\end{aligned}
$$

h_{ij} は Σ^{-1} の i, j 要素
正規分布はマハラノビス距離の指数関数

$$
N_x(M,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}(X-M)^T\Sigma^{-1}(X-M))
$$

パラメータ推定

$\mathbf{y}$ を $\mathbf{x}_1, \ldots, \mathbf{x}_n$ の関数とする:

$$\mathbf{y} = f(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

期待値 m_y と分散 σ_y^2:

$$m_y = E\{\mathbf{y}\} \text{ and } \sigma_y = Var\{\mathbf{y}\}.$$

平均 M や共分散行列 Σ (の要素) は $\mathbf{y}$ の例
特に、$\mathbf{y} = \mathbf{x}_1^{i_1} \cdots \mathbf{x}_n^{i_n}$ の時、m_y は $(i_1 + \cdots + i_n)$ 次モーメントと呼ばれる

推定値のモーメント

標本平均:

$$\hat{m}_y = \frac{1}{N}\sum_{k=1}^{N} y_k.$$

標本平均 $\hat{m}_y$ は 不偏性 *(unbiased)* を有する:

$$\begin{aligned} E\{\hat{\mathbf{m}}_y\} &= \frac{1}{N}\sum_{k=1}^{N} E\{\mathbf{y}_k\} \\ &= \frac{1}{N}\sum_{k=1}^{N} m_y = m_y. \end{aligned}$$

推定値のモーメント

標本平均の分散:

$$
\begin{aligned}
Var\{\hat{\mathbf{m}}_y\} &= E\{(\hat{\mathbf{m}}_y - m_y)^2\} \\
&= \frac{1}{N^2}\sum_{k=1}^{N}\sum_{\ell=1}^{N} E\{(\mathbf{y}_k - m_y)(\mathbf{y}_\ell - m_y)\} \\
&= \frac{1}{N^2}\sum_{k=1}^{N} E\{(\mathbf{y}_k - m_y)^2\} = \frac{1}{N}\sigma_y^2.
\end{aligned}
$$

N が無限になると $Var\{\hat{\mathbf{m}}_y\}$ は 0 に収束するので、標本平均は 一致性 *(consistency)* を有する

中心モーメント

分散や共分散のような中心モーメントはより複雑になる

$$\mathbf{y} = (\mathbf{x}_i - m_i)(\mathbf{x}_j - m_j)$$

m_i や m_j は未知なので、標本推定で置き換える

$$\mathbf{y} = (\mathbf{x}_i - \hat{\mathbf{m}}_i)(\mathbf{x}_j - \hat{\mathbf{m}}_j)$$

標本共分散行列

$$
\begin{aligned}
\hat{\Sigma} &= \frac{1}{N}\sum_{k=1}^{N}(\mathbf{X}_k - \hat{\mathbf{M}})(\mathbf{X}_k - \hat{\mathbf{M}})^T \\
&= \frac{1}{N}\sum_{k=1}^{N}\{(\mathbf{X}_k - M) - (\hat{\mathbf{M}} - M)\}\{(\mathbf{X}_k - M) - (\hat{\mathbf{M}} - M)\}^T \\
&= \frac{1}{N}\sum_{k=1}^{N}(\mathbf{X}_k - M)(\mathbf{X}_k - M)^T - (\hat{\mathbf{M}} - M)(\hat{\mathbf{M}} - M)^T \\
E\{\hat{\Sigma}\} &= \Sigma - E\{(\hat{\mathbf{M}} - M)(\hat{\mathbf{M}} - M)^T\} \\
&= \Sigma - \frac{1}{N}\Sigma = \frac{N-1}{N}\Sigma
\end{aligned}
$$

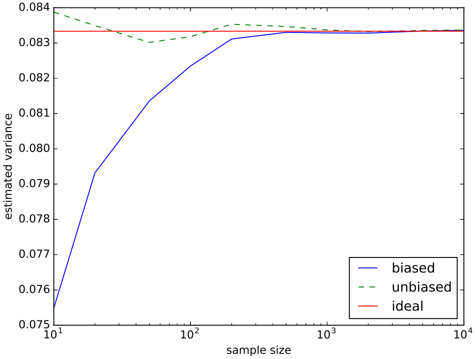
標本共分散行列

従って不偏ではない。不偏推定は以下の通り

$$\hat{\Sigma} = \frac{1}{N-1}\sum_{k=1}^{N}(\mathbf{X}_k - \hat{\mathbf{M}})(\mathbf{X}_k - \hat{\mathbf{M}})^T.$$

演習

biasvar.py (biasvar.m, biasvar.sci)
一様分布から異なる標本数の標本を得て、偏った推定量並びに不偏推定量の分散を計算せよ。
N (=1000) 回計算し、その平均値をプロットせよ。



線形変換

ベクトル $\mathbf{X}$ を行列 A により別のベクトル $\mathbf{Y}$ に線形変換することを考える

$$\mathbf{Y} = A^T\mathbf{X}$$

$\mathbf{Y}$ の平均ベクトルと共分散行列は以下の通り

$$\begin{aligned}
M_Y &= E\{\mathbf{Y}\} = A^T E\{\mathbf{X}\} = A^T M_X \\
\Sigma_Y &= E\{(\mathbf{Y} - M_Y)(\mathbf{Y} - M_Y)^T\} \\
&= A^T E\{(\mathbf{X} - M_X)(\mathbf{X} - M_X)^T\} A \\
&= A^T \Sigma_X A.
\end{aligned}$$

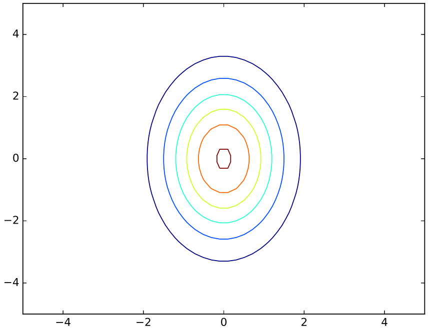
Python sample (linearnorm.py)

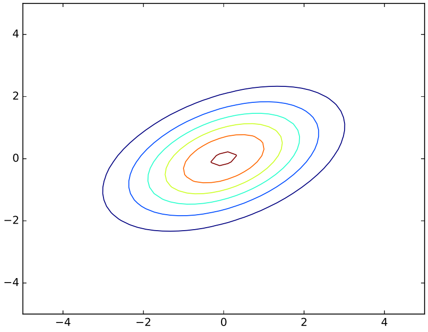
```
import numpy as np
import matplotlib.pyplot as plt

cov = np.matrix([[1, 0], [0, 3]])
icov = np.linalg.inv(cov)
r = np.matrix([[np.cos(np.pi / 3), -np.sin(np.pi / 3)],
              [np.sin(np.pi / 3), np.cos(np.pi / 3)]])
xx, yy = np.meshgrid(np.linspace(-5, 5), np.linspace(-5, 5))
plt.figure()
plt.axis('equal')
p = 1. / (2. * np.pi * np.sqrt(np.linalg.det(cov))) \
    * np.exp(-1. / 2. * \
             (icov[0, 0] * xx * xx \
              + (icov[0, 1] + icov[1, 0]) * xx * yy \
              + icov[1, 1] * yy * yy))
plt.contour(xx, yy, p)
```

Python sample (linearnorm.py)

```
# plt.savefig('linearnorm-1.eps')
plt.figure()
plt.axis('equal')
cov2 = r.T.dot(cov).dot(r)
icov2 = np.linalg.inv(cov2)
pp = 1. / (2. * np.pi * np.sqrt(np.linalg.det(cov2))) \
    * np.exp(-1. / 2. \
             * (icov2[0, 0] * xx * xx \
                + (icov2[0, 1] + icov2[1, 0]) * xx * yy \
                + icov2[1, 1] * yy * yy))
plt.contour(xx, yy, pp)
# plt.savefig('linearnorm-2.eps')
plt.show()
```

Python sample (linearnorm.py)





Matlab sample

```
cov=[1 0; 0 3];
r=[cos(pi/3) -sin(pi/3); sin(pi/3) cos(pi/3)];
[xx,yy]=meshgrid(-5:0.5:5,-5:0.5:5);
xy=[xx(:) yy(:)];
figure
p=1/(2*pi*sqrt(det(cov))) * ...
    exp(-1/2*diag(xy*inv(cov)*xy'));
contour(xx,yy,reshape(p,size(xx)),10);
figure
p=1/(2*pi*sqrt(det(cov))) * ...
    exp(-1/2*diag(xy*inv(r'*cov*r)*xy'));
contour(xx,yy,reshape(p,size(xx)),10);
```

マハラノビス距離の不変性

$$
\begin{aligned}
d_Y^2(Y) &= (Y - M_Y)^T \Sigma_Y^{-1} (Y - M_Y) \\
&= (X - M_X)^T A A^{-1} \Sigma_X^{-1} (A^T)^{-1} A^T (X - M_X) \\
&= (X - M_X) \Sigma_X^{-1} (X - M_X) \\
&= d_X^2(X).
\end{aligned}
$$

正規直交変換

平均が 0 となるように座標系を変換しよう:

$$Z = X - M$$

これにより

$$d_Z^2(Z) = Z^T \Sigma^{-1} Z.$$

$Z^T Z = 1$ (Z は単位ベクトル) という条件のもと $d_Z^2(Z)$ を最大化するベクトル Z を求めよう

$$\frac{\partial}{\partial Z}\{Z^T \Sigma^{-1} Z - \mu(Z^T Z - 1)\} = 2\Sigma^{-1} Z - 2\mu Z = 0. \text{ (導出してみよ)}$$

ラグランジュ未定乗数決定法

以下の最適化問題を考える

$$\text{maximize } f(x) \text{ subject to } g(x) = 0$$

ラグランジュ未定乗数 λ を用い、以下のラグランジュ関数を考える

$$\Lambda(x, \lambda) = f(x) + \lambda g(x)$$

最初の最適化問題は、以下により解くことができる

$$\frac{\partial}{\partial x}\Lambda = \frac{\partial}{\partial \lambda}\Lambda = 0$$

正規直交変換

$\Sigma^{-1}Z = \mu Z$ or $\Sigma Z = \lambda Z$ ($\lambda = \frac{1}{\mu}$) は固有値問題として解くことができる
特性方程式:

$$|\Sigma - \lambda I| = 0.$$

この方程式を満たす λ を 固有値, 対応する Z を 固有ベクトルと呼ぶ
対称 $n \times n$ 行列 Σ については、n 個の実固有値 $\lambda_1, \ldots, \lambda_n$ と n 個の実固有ベクトル $\phi_1, \ldots, \phi_n$ が得られる
異なる固有値に対応する固有ベクトルは互いに直行する

$$\phi_i^T \phi_j = 0 \;\; i \neq j$$

$\Phi = [\phi_1 \cdots \phi_n]$ を固有ベクトル行列と呼ぶ

正規直交変換

明らかに、

$$\Sigma\Phi = \Phi\Lambda$$
$$\Phi^T\Phi = I$$

正規直交変換

Φ を変換行列としよう

$$Y = \Phi^T X$$

これにより

$$\Sigma_Y = \Phi^T \Sigma_X \Phi = \Lambda.$$

特性:

- Y は X を $\phi_1, \ldots, \phi_n$ で張られる新たな座標系で表している
- 共分散行列は対角化される。これにより対応する確率変数は一般には無相関となり、正規分布の場合は独立となる
- 固有ベクトルは $d_Z^2(Z)$ を最大化するので、この変換は分散の主成分を選ぶことになる
- この変換を 正規直行変換 と呼ぶ

白色化

正規直行変換を施したのち、さらに別の変換 $\Lambda^{-\frac{1}{2}}$ を施すことにより、共分散行列を単位行列に変換することができる

$$Y = \Lambda^{-\frac{1}{2}}\Phi^T X = (\Phi\Lambda^{-\frac{1}{2}})^T X$$

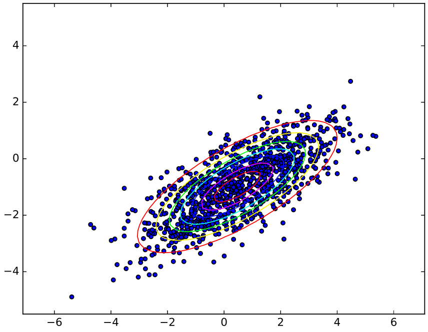
$$\Sigma_Y = \Lambda^{-\frac{1}{2}}\Phi^T\Sigma_X\Phi\Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}} = I.$$

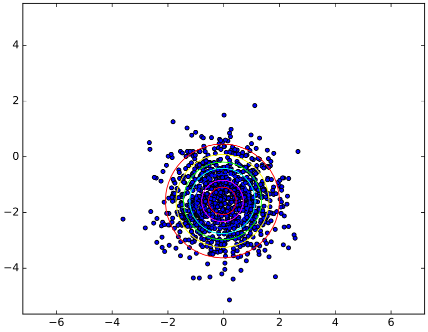
$\Phi\Lambda^{-\frac{1}{2}}$ を白色化と呼ぶ
特性:

- 白色化は正規直行変換ではない
- 白色化後は、共分散行列はいかなる正規直行変換に対しても不変である

$$\Psi^T I \Psi = \Psi^T\Psi = I.$$

Whitening Transformation (whitening.py)





同時対角化

二つの異なる対称行列 Σ_1 と Σ_2 とを線形変換により同時に対角化できる

(1) Σ_1 を以下により白色化

$$Y = \Theta^{-\frac{1}{2}} \Phi^T X$$

ただし Θ と Φ は Σ_1 の固有値行列と固有ベクトル行列

$$\Sigma_1 \Phi = \Phi \Theta \text{ かつ } \Phi^T \Phi = I.$$

これにより Σ_1 と Σ_2 は以下のように変換される

$$\Theta^{-\frac{1}{2}} \Phi^T \Sigma_1 \Phi \Theta^{-\frac{1}{2}} = I$$

$$\Theta^{-\frac{1}{2}} \Phi^T \Sigma_2 \Phi \Theta^{-\frac{1}{2}} = K$$

一般に K は対角行列ではない

(2) 正規直行変換により K を対角化すればよい

$$Z = \Psi^T Y$$

ただし Ψ と Λ は K の固有ベクトル行列と固有値行列

$$K\Psi = \Psi\Lambda \text{ and } \Psi^T\Psi = I.$$

よって Ψ により

$$\Psi^T I \Psi = \Psi^T \Psi = I$$
$$\Psi^T K \Psi = \Lambda$$

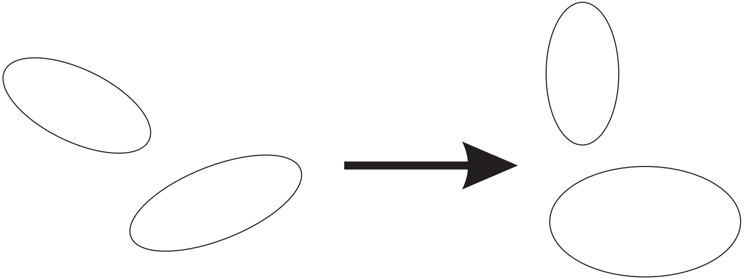
こうして両行列とも対角化できた

宿題

- 今回は宿題としてプログラミング課題と非プログラミング課題を課する
- プログラミング課題か非プログラミング課題かのいずれかを解いて提出すること
- プログラミング課題を解くことを推奨する
- もちろん両方解いてもらえると嬉しい

プログラミング課題

1. 二つのデータファイル (data.mat と data2.mat) が与えられている。それぞれ二つの正規分布に従い発生させた点群 (2D 2000 点) と、正規分布の十分統計量を含んでいる。同時対角化を実装し、与えられた正規分布のパラメータに応じて点群に適用せよ。それぞれのファイルについて同時対角化の前と後の散布図を図示せよ。



2. (余力問題) 同時対角化を適用し、散布図に正規分布の等高線を同時に図示せよ。

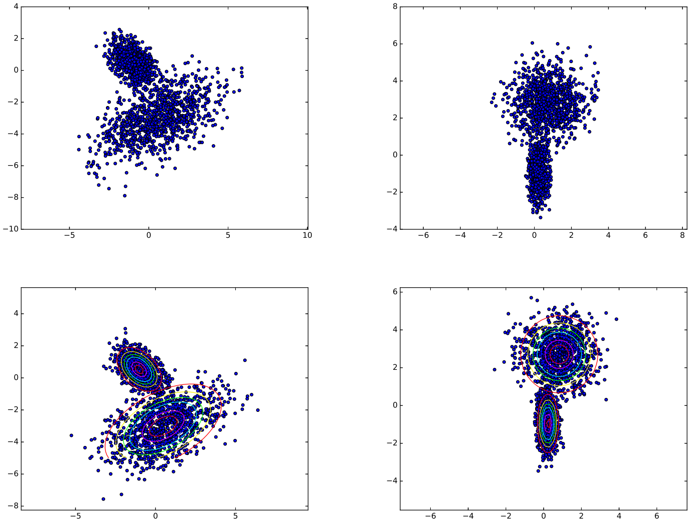
プログラミング課題

データは以下のように読み込み可能:

```
from scipy.io import loadmat
data = loadmat('data.mat')
x = data['x']  # data matrix
cov1 = data['cov1']  # covariance matrix 1
cov2 = data['cov2']  # covariance matrix 2
m1 = data['m1']  # mean vector 1
m2 = data['m2']  # mean vector 2
```

ファイルは講義の Web ページよりダウンロード可能。
Google Colaboratory を使う場合は、Google drive 上の共有フォルダから直接読み込み可能。
方法は講義の Web ページの colab を参照のこと。Colaboratory を使う場合でも、ダウンロードしたファイルを自分のドライブに展開して読み込み可能。

プログラミング課題


4 2 0 −2 −4 −6 −8 −10
−5 0 5 10
8 6 4 2 0 −2 −4
−6 −4 −2 0 2 4 6 8
4 2 0 −2 −4 −6 −8
−5 0 5
6 4 2 0 −2 −4
−6 −4 −2 0 2 4 6

非プログラミング課題

(1) $(\mathbf{x}_1, \mathbf{x}_2, \cdots)$ が多次元正規分布に従うとする。$\mathbf{x}_i$ と $\mathbf{x}_j (i \neq j)$ が無相関ならば独立でもあることを示せ。

(2) $\mathbf{x} \sim N_x(0, \sigma)$ と $\mathbf{y} = \mathbf{x}^2$ とは無相関だが独立ではないことを示せ。

宿題の提出方法

- PDF ファイルとして提出せよ。
- プログラミング課題については、プログラム本体、プログラムの簡単な説明、結果、結果の説明を記載した PDF を準備せよ。プログラムリストのみのレポートとしないように！
- 通常締め切りは二回後の講義の日とする (例えば今日の宿題の締め切りは 5 月 2 日)
- ITC-LMS にアップロードせよ。メイルで送らないように！