

参考書:

- [1] 長尾真:「画像認識論」, コロナ社, 1983年.
- [2] 鳥脇純一郎:「認識工学」, コロナ社, 1993年.
- [3] 石井, 上田, 前田, 村瀬:「わかりやすいパターン認識」, オーム社, 1998年.
- [4] K. Fukunaga: “*Introduction to Statistical Pattern Recognition*,” Second Edition, Academic Press, 1990.
- [5] C. M. Bishop: “*Neural Networks for Pattern Recognition*,” Oxford University Press, 1995.
- [6] 金谷健一:「これなら分かる応用数学教室
—最小二乗法からウェーブレットまで—」,
共立出版, 2003年.

講義ノート: <http://cis.k.hosei.ac.jp/~wakahara>

講義の進め方:

1. パターン認識の理論的基礎を理解する
→ 計算問題を解く
2. 特に, 統計的／確率論的な取り扱いになじむようにする
→ 線形代数, 確率論が大切
3. 質疑を活発に行う

評価方法:

出席点 20点 宿題点 40点
期末試験 40点

第1回講義「パターン認識」

数学的基礎

[A.1] 線形代数

[A.2] 確率論

A.1 線形代数

A.1.1 記号と予備知識

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

$$\mathbf{x}^t = (x_1 \ x_2 \ \dots \ x_d)$$

ベクトルと行列の
定義は基本中の
基本

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & \dots & m_{1d} \\ m_{21} & m_{22} & m_{23} & \dots & m_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & m_{n3} & \dots & m_{nd} \end{pmatrix}$$

$$\mathbf{M}^t = \begin{pmatrix} m_{11} & m_{21} & \dots & m_{n1} \\ m_{12} & m_{22} & \dots & m_{n2} \\ m_{13} & m_{23} & \dots & m_{n3} \\ \vdots & \vdots & \ddots & \vdots \\ m_{1d} & m_{2d} & \dots & m_{nd} \end{pmatrix}$$

・ 非負行列 $m_{ij} \geq 0, \forall i, j$

・ $(d \times d)$ 正方行列 $I = (\delta_{ij})$: 単位行列

$m_{ij} = m_{ji}$: 対称行列

$m_{ij} = -m_{ji}$: 反対称行列

・ 行列とベクトルの積: $M \mathbf{x} = \mathbf{y}$

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} & \dots & m_{1d} \\ m_{21} & m_{22} & m_{23} & \dots & m_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & m_{n3} & \dots & m_{nd} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

行列とベクトルの
積も基本中の
基本

$$y_i = \sum_{j=1}^d m_{ij} x_j \quad (1 \leq i \leq n)$$

A.1.2 内積

$$\mathbf{x}^t \mathbf{y} = \sum_{i=1}^d x_i y_i = \mathbf{y}^t \mathbf{x} \quad \mathbf{x} \cdot \mathbf{y}, (\mathbf{x}, \mathbf{y}) \quad \text{とも記す}$$

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^t \mathbf{x}} \quad \text{: ユークリッドノルム} \quad \text{内積はしっかり覚えること}$$

2つの d 次元ベクトルのなす角 θ

$$\cos \theta = \frac{\mathbf{x}^t \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad \mathbf{x}^t \mathbf{y} = 0 \quad \text{: 直交}$$
$$|\mathbf{x}^t \mathbf{y}| = \|\mathbf{x}\| \|\mathbf{y}\| \quad \text{: 平行}$$

$$|\mathbf{x}^t \mathbf{y}| \leq \|\mathbf{x}\| \|\mathbf{y}\| \quad \text{: Cauchy-Schwarz の不等式}$$

この不等式は有用

A.1.3 外積

$$\mathbf{M} = \mathbf{x} \mathbf{y}^t = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} (y_1 \ y_2 \ \dots \ y_n) = \begin{pmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \dots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_d y_1 & x_d y_2 & \dots & x_d y_n \end{pmatrix}$$

A.1.4 線形独立

ベクトルの集合があり, どのベクトルも他のベクトルの線形結合として表せない場合に線形独立と呼ぶ. d 個の線形独立なベクトルは d 次元のベクトル空間を張り, その空間内の任意のベクトルはそれら d 個のベクトルの線形結合で表すことができる.

A.1.5 行列の微分

- ・ スカラー関数の微分

$$\nabla f(\mathbf{x}) = \text{grad } f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} =$$

$$\begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_d} \end{pmatrix}$$

∇はナブラと読みます

- ・ 行列 M と y が x に独立であるとき

$$\frac{\partial}{\partial \mathbf{x}} [M \mathbf{x}] = M \quad \frac{\partial}{\partial \mathbf{x}} [y^t \mathbf{x}] = \frac{\partial}{\partial \mathbf{x}} [\mathbf{x}^t y] = y$$

$$\frac{\partial}{\partial \mathbf{x}} [\mathbf{x}^t M \mathbf{x}] = [M + M^t] \mathbf{x}$$

A.1.6 行列式とトレース

- ・ $(d \times d)$ 正方行列の行列式はスカラー: $|M|$

$$i \begin{pmatrix} & & & j & & \\ m_{11} & m_{12} & \dots & \otimes & \dots & m_{1d} \\ m_{21} & m_{22} & \dots & \otimes & \dots & m_{2d} \\ & & & \otimes & & \\ & & & \otimes & & \\ \otimes & \otimes & \dots & \otimes & \dots & \otimes \\ & & & \otimes & & \\ m_{d1} & m_{d2} & \dots & \otimes & \dots & m_{dd} \end{pmatrix} = M_{i|j}$$

M の第 i 行と
第 j 列を取り
除いた行列

- $|M|$ は再帰的に次式で計算できる.

$$|M| = m_{11}|M_{1|1}| - m_{21}|M_{2|1}| + m_{31}|M_{3|1}| - \dots \pm m_{d1}|M_{d|1}|$$

$$d = 1 \quad |M| = m_{11}$$

$$d = 2 \quad |M| = m_{11}m_{22} - m_{21}m_{12}$$

- 性質 $|M| = |M^t| \quad |MN| = |M||N|$

$$M \text{ の列ベクトルが線形独立でない} \rightarrow |M| = 0$$

- トレース

$$\text{tr}[M] = \sum_{i=1}^d m_{ii}$$

$|M|$ および $\text{tr}[M]$ は座標軸の回転に関して不変である.

A.1.7 逆行列

- ・ 行列式が 0 でなければ逆行列 M^{-1} が存在する.

$$M M^{-1} = M^{-1} M = I \quad \text{逆行列の定義は重要}$$

$$M^{-1} = \frac{\tilde{M}}{|M|} \quad \tilde{M} : M \text{ の随伴行列}$$
$$\tilde{m}_{ij} = (-1)^{i+j} |M_{j|i}|$$

- ・ M が正方でないときは, 次式の擬似逆行列を用いる.

$$M^+ M = I \quad M^+ = [M^t M]^{-1} M^t$$

A.1.8 固有ベクトルと固有値

- ・ $(d \times d)$ 正方行列 M に対して次の x に関する1次方程式を解く.

$$Mx = \lambda x \quad \rightarrow \quad (M - \lambda I)x = 0$$

$$|M - \lambda I| = \lambda^d + a_1 \lambda^{d-1} + \dots + a_{d-1} \lambda + a_d = 0$$

- ・ 実対称行列の固有値 $\{\lambda_1, \dots, \lambda_d\}$ はすべて実数であり、対応する固有ベクトル $\{e_1, \dots, e_d\}$ も実数ベクトルとなる.

$$Me_j = \lambda_j e_j \quad e_i^t e_j = \delta_{ij}, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$$

$$\text{tr}[M] = \sum_{i=1}^d \lambda_i \quad |M| = \prod_{i=1}^d \lambda_i$$

固有値, 固有ベクトルの意味を理解する

A.2 確率論

A.2.1 離散的確率変数

$$x \in X = \{v_1, v_2, \dots, v_m\}$$

$$p_i = \Pr[x = v_i], \quad (i = 1, \dots, m)$$

$$p_i \geq 0 \quad \text{and} \quad \sum_{i=1}^m p_i = 1$$

- ・ 確率の集合 $\{p_1, p_2, \dots, p_m\}$ を確率質量関数 $P(x)$ を使って書くとわかり易い.

$$P(x) \geq 0 \quad \text{and} \quad \sum_{x \in X} P(x) = 1$$

A.2.2 期待値

- 確率変数 x の期待値 (平均値):

$$E[x] = \mu = \sum_{x \in X} x P(x) = \sum_{i=1}^m v_i p_i$$

期待値の操作 $E[\]$ の意味を理解すること

一般に $E[f(x)] = \sum_{x \in X} f(x) P(x)$ E は線形演算子であることに注意

- 確率変数 x の分散

平均値と分散の定義は暗記

$$\text{Var}[x] = \sigma^2 = E[(x - \mu)^2] = \sum_{x \in X} (x - \mu)^2 P(x)$$

一般に $\text{Var}[f(x)] = \sum_{x \in X} (f(x) - E[f(x)])^2 P(x)$

A.2.3 離散的確率変数の対

$$x \in X = \{v_1, v_2, \dots, v_m\}, \quad y \in Y = \{w_1, w_2, \dots, w_n\}$$

$$P(x, y) \geq 0 \quad \text{and} \quad \sum_{x \in X} \sum_{y \in Y} P(x, y) = 1$$

- ・ 結合確率質量関数 $P(x, y)$ の周辺分布

$$P_x(x) = \sum_{y \in Y} P(x, y), \quad P_y(y) = \sum_{x \in X} P(x, y)$$

- ・ 統計的独立性

$$P(x, y) = P_x(x) P_y(y)$$

A.2.4 2変量関数の期待値

$$E[f(x, y)] = \sum_{x \in X} \sum_{y \in Y} f(x, y) P(x, y)$$

$$\mu_x = E[x] = \sum_{x \in X} \sum_{y \in Y} x P(x, y) = \sum_{x \in X} x P_x(x)$$

$$\mu_y = E[y] = \sum_{x \in X} \sum_{y \in Y} y P(x, y) = \sum_{y \in Y} y P_y(y)$$

$$\sigma_x^2 = \text{Var}[x] = E[(x - \mu_x)^2] = \sum_{x \in X} (x - \mu_x)^2 P_x(x)$$

$$\sigma_y^2 = \text{Var}[y] = E[(y - \mu_y)^2] = \sum_{y \in Y} (y - \mu_y)^2 P_y(y)$$

- ・ 相互モーメント

$$\begin{aligned}\sigma_{xy} &= E[(x - \mu_x)(y - \mu_y)] \\ &= \sum_{x \in X} \sum_{y \in Y} (x - \mu_x)(y - \mu_y) P(x, y)\end{aligned}$$

- ・ $\lambda x + y$ (λ : 実数) の分散が負にならないことから次の不等式が導き出される.

$$\sigma_{xy}^2 \leq \sigma_x^2 \sigma_y^2 \quad (1)$$

- ・ 相関係数

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \quad -1 \leq \rho \leq 1$$

$\rho = 1$: 最大の正相関
 $\rho = 0$: 無相関
 $\rho = -1$: 最大の負相関

A.2.5 総確率の法則とベイズ則

事象 A が相互に排他的な子事象 $\{A_i\}$ に分けられたとすると A の起こる確率は子事象 A_i の起こる確率の総和となる.

$$P(A) = \sum_i P(A_i) \quad : \text{総確率の法則}$$

したがって

$$P(y) = \sum_{x \in X} P(x, y), \quad P(x) = \sum_{y \in Y} P(x, y) \quad (2)$$

一方, 条件付き確率の定義より

$$P(x, y) = P(y | x) P(x) = P(x | y) P(y) \quad (3)$$

条件付き確率の定義は覚える

式(2)と式(3)を組み合わせると次式が得られる.

$$P(x | y) = \frac{P(y | x) P(x)}{\sum_{x \in X} P(y | x) P(x)} \quad \text{: ベイズ則}$$

ベイズ則は覚えること

ベイズ則の意義:

x を原因, y を結果と考える. 結果 y を知って原因 x を推定する際の条件付き確率 $P(x | y)$ を計算する手段を与えている.
すなわち, 比較的求めやすい

[1] y を観測する前の x の事前確率 $P(x)$, および

[2] x という原因があるときの y という結果の起こる
条件付き確率 $P(y | x)$

を用いることにより, 事後確率 $P(x | y)$ が算出可能となる.

A.2.6 ベクトル確率変数

- ・ d 次元ベクトルの確率変数 \mathbf{x} を扱う.

$$\boldsymbol{\mu} = E[\mathbf{x}] = \sum_{\mathbf{x}} \mathbf{x} P(\mathbf{x}) \quad : \text{平均ベクトル}$$

$$\boldsymbol{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t] \quad : \text{共分散行列}$$

$\boldsymbol{\Sigma}$ は対称行列で, 任意の d 次元ベクトル \mathbf{w} に対して

$$\mathbf{w}^t \boldsymbol{\Sigma} \mathbf{w} = E[((\mathbf{x} - \boldsymbol{\mu})^t \mathbf{w})^2] \geq 0$$

を満たすことより, 半正定値の行列である. これより, 共分散行列 $\boldsymbol{\Sigma}$ の固有値はすべて非負であることが示される.

A.2.7 連続な確率変数と正規分布

- ・ 確率質量関数 $P(x)$ の代わりに確率密度関数 $p(x)$ を考える.

$$\Pr [x \in (a, b)] = \int_a^b p(x) dx$$

- ・ 正規分布 $N(\mu, \sigma^2)$

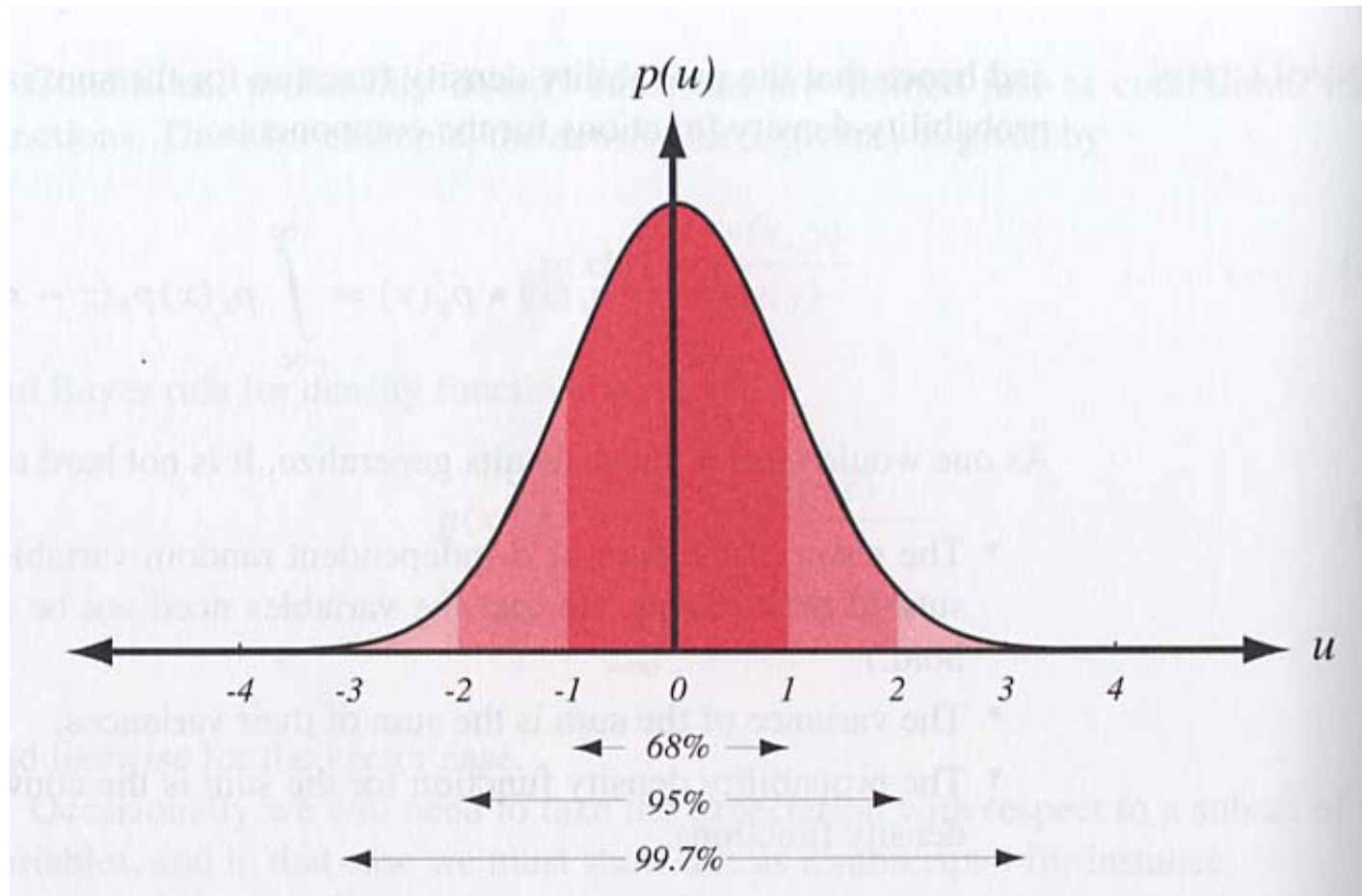
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

1次元正規分布の式は暗記する

$$E[1] = \int_{-\infty}^{\infty} p(x) dx = 1 \quad E[x] = \int_{-\infty}^{\infty} x p(x) dx = \mu$$

$$E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx = \sigma^2$$

1次元正規分布 $p(u) \sim N(0, 1)$



$|u| \leq 1$ の範囲に 68 % が, $|u| \leq 2$ の範囲には 95 %,
 $|u| \leq 3$ の範囲には 99.7 % の確率質量が存在する.

練習問題1-1

σ_x , σ_y および σ_{xy} の定義式を用いて, $\lambda x + y$ (λ : 実数) の分散:

$$\text{Var}[\lambda x + y] = E[(\lambda x + y) - E[\lambda x + y]]^2]$$

が負にならないことから, 次の不等式を導きなさい.

$$\sigma_{xy}^2 \leq \sigma_x^2 \sigma_y^2$$

ヒント: $E[\lambda x + y] = \lambda E[x] + E[y] = \lambda \mu_x + \mu_y$ 等を用いて $\text{Var}[\lambda x + y]$ を変形する. 実数 λ に関する2次式が非負となる条件には2次方程式の判別式を用いなさい.

練習問題1-2

2つの箱がある. 第1の箱には, リンゴが7個, オレンジが4個入っている. 第2の箱には, リンゴが6個, オレンジが9個入っている. このとき, ランダム(等確率)にどちらかの箱を選択して, 中から果物を1つ取り出すものとする.

今, 箱を選んで果物を取り出したら, リンゴであった. 選んだ箱が第1の箱である確率を計算しなさい.

ヒント: ベイズ則を用いなさい. 原因 x が箱の選択, 結果 y が取り出した果物の種類, に対応するとして, $P(x | y)$ を計算しなさい.

課題の提出要領

【提出方法】

- ・A4用紙に手書きもしくはMS Wordで作成
- ・複数枚の場合はホッチキス止め
- ・第1頁の冒頭に

パターン認識 第〇回課題レポート

提出年月日 2012年△月□日

学籍番号 氏名

を必ず記す

【提出期限】

- ・次回講義の冒頭, 遅刻は認めず