線形代数

- ベクトル：数を１列に並べたもの
- 行列：数を長方形状に並べたもの
- 線形性：関数 f が線形とは
  1. $f(x)+f(y)=f(x+y)$
  2. $f(ax)=af(x)$

  が成り立つこと
- 行列の行と列の個数の組を「型」又は「サイズ」という
- 行列の演算
  - 和（差）
    - 型が同じ行列でなければ演算できない
    - 対応する成分同士で演算する
  - 積
    - 行列のスカラ倍は全ての成分をスカラ倍する
    - 行列同士の積はサイズが(l, m) と (m, n)のように，１つ目の列と２つ目の行の個数が等しい場合にのみ演算が可能で，演算結果のサイズは(l, n)になる
- 特殊な行列
  - 零行列：全ての成分が 0 の行列
  - 転置行列：ある行列の行と列を入れ替えた行列
  - 正方行列：行と列の個数が同じ行列
    - 対角行列：行番号と列番号が等しい成分を対角成分といい，対角成分以外の成分が 0 である行列
    - スカラー行列：対角成分が全て同じ値である対角行列
    - 単位行列：対角成分が全て 1 である対角行列
    - 対称行列：転置行列と元の行列が等しい行列 $\boldsymbol{X}^{\boldsymbol{T}}=\boldsymbol{X}$
    - 交代行列：転置行列が元の行列の-1 倍となる行列 $\boldsymbol{X}^{\boldsymbol{T}}=-\boldsymbol{X}$
    - 逆行列：ある行列との積が単位行列となるような行列 $\boldsymbol{X}\boldsymbol{X}^{-1}=\boldsymbol{X}^{-1}\boldsymbol{X}=\boldsymbol{E}$
    - 正則行列：逆行列を持つ正方行列
- 連立 1 次方程式の解法
  - 掃き出し法：係数行列に関する基本変形（行の入れ替え・行同士の加減・行の定数倍）により求める
  - 掃き出し法は基本変形を表す行列の積で表すことができる
- 逆行列の存在に必要な条件は正方行列の行列式が 0 ではないこと
  - n 次正方行列 A の行列式

    $$det(\boldsymbol{A})=\sum_{\sigma\in S_n}(sgn\,\sigma)a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}$$

    平たく言うと，n 個の成分を行と列が重複しないように選んで積をとる．これを全ての成分を重複なく行う．これにより作られた値の和が行列式となる．ただし，選択した成分の位置により符号 $(sgn\,\sigma)$ の値が変わる（位置の入れ替えが偶数で+1，奇数で-1）
  - 2 次の正方行列は $det(\boldsymbol{A})=a_{11}a_{22}-a_{12}a_{21}$
  - 3 次の正方行列はサラスの方法
  - 4 次以上の正方行列は余因子展開
    - 余因子展開：正方行列において，1 つの行又は列を基準とし，基準のうちの 1 つの成分を取り出して係数とし，行列から取り出した成分が位置している行と列を除いた行列に成分の位置による符号をかけたもの（これを余因子という）を掛ける．これを基準としたところの全成分に行い和をとる．
  - 行列式の特徴（列ベクトル・行ベクトルどちらでも成り立つ）
    - $|\boldsymbol{a_1}\,\boldsymbol{a_2}\,\boldsymbol{b}+\boldsymbol{c}\,\boldsymbol{a_4}|=|\boldsymbol{a_1}\,\boldsymbol{a_2}\,\boldsymbol{b}\,\boldsymbol{a_4}|+|\boldsymbol{a_1}\,\boldsymbol{a_2}\,\boldsymbol{c}\,\boldsymbol{a_4}|$ （列の和は行列式の和）
    - $|\boldsymbol{a_1}\,\boldsymbol{a_2}\,c\,\boldsymbol{a_3}\,\boldsymbol{a_4}|=c|\boldsymbol{a_1}\,\boldsymbol{a_2}\,\boldsymbol{a_3}\,\boldsymbol{a_4}|$ （列をスカラ倍すると行列式もスカラ倍）
    - $|\boldsymbol{a_1}\,\boldsymbol{a_3}\,\boldsymbol{a_2}\,\boldsymbol{a_4}|=-|\boldsymbol{a_1}\,\boldsymbol{a_2}\,\boldsymbol{a_3}\,\boldsymbol{a_4}|$ （列を入れ替えると符号が変わる）
    - 2 つの列が等しい正方行列の行列式は 0
- 固有値・固有ベクトル

  零ベクトルではないベクトル $\boldsymbol{x}$ を正方行列 A で変換させたとき，元のベクトルのスカラ倍となる．すなわち， $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$ を満たすとき λ を A の固有値， $\boldsymbol{x}$ を A の固有ベクトルという．

    - 固有値の求め方：固有方程式の解を求める
      $\phi_A(\lambda)=|\lambda \boldsymbol{E}-\boldsymbol{A}|=0$
    - 固有ベクトルの求め方：各固有値について以下の連立方程式を解く
      $(\lambda \boldsymbol{E}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$
        - 重解のときはベクトルの和で表現される
- 固有値分解
  固有値と固有ベクトルの対応に注意してそれぞれ並べて行列 $\boldsymbol{\Lambda}, \boldsymbol{X}$ を作る($\boldsymbol{\Lambda}$ は固有値を対角成分とする対角行列）．すると $\boldsymbol{A}\boldsymbol{X}=\boldsymbol{X}\boldsymbol{\Lambda}$ が成り立ち， $\boldsymbol{A}=\boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{-1}$ と変形できる
    - 行列の累乗の計算
      $\boldsymbol{A}^n=\underbrace{\boldsymbol{A}\boldsymbol{A}\cdots\boldsymbol{A}}_{\times n}=\underbrace{\boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{-1}\boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{-1}\cdots\boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{-1}}_{\times n}=\boldsymbol{X}\boldsymbol{\Lambda}^n\boldsymbol{X}^{-1}$ $\boldsymbol{\Lambda}^n$ は，固有値の n 乗が成分の対角行列
- 特異値分解
  固有値分解は正方行列にのみ適用可能であり，これを正方行列以外に拡張する
    - 特異値：以下の条件を満たす正の数σ
        - 零行列ではない任意の m×n 行列 A に対し，
            - $\boldsymbol{A}\boldsymbol{v}=\sigma\boldsymbol{u}$
            - $\boldsymbol{A}^T\boldsymbol{u}=\sigma\boldsymbol{v} \quad (\boldsymbol{u}\neq\boldsymbol{0}, \boldsymbol{v}\neq\boldsymbol{0})$
  ここで，第１式の両辺に左から$\boldsymbol{A}$を，第２式の両辺に左から$\boldsymbol{A}^T$をかけると
            - $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}=\sigma\boldsymbol{A}^T\boldsymbol{u}=\sigma^2\boldsymbol{v}$
            - $\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{u}=\sigma\boldsymbol{A}\boldsymbol{v}=\sigma^2\boldsymbol{u}$
  となり，$\boldsymbol{A}\boldsymbol{A}^T$と$\boldsymbol{A}^T\boldsymbol{A}$の固有値分解に帰着される．以上より固有ベクトル u を並べた行列 U と固有ベクトル v を並べた行列 V，それに固有値 σ^2 の対角行列を$\boldsymbol{D}\boldsymbol{D}^T, \boldsymbol{D}^T\boldsymbol{D}$とすると
            - $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}=\boldsymbol{V}\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{V}^{-1}=\boldsymbol{V}\boldsymbol{D}^T\boldsymbol{U}^{-1}\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{-1}$
            - $\boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{D}^T\boldsymbol{U}^{-1}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{U}^{-1}$
  以上より，$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{-1}$ （D は m×n 行列で対角成分以外は 0）

【参考文献】

1. 藤岡 敦『手を動かしてまなぶ線形代数』裳華房　2015.11
2. 平井 有三『はじめてのパターン認識』森北出版　2012.07
3. Ian Goodfellow ら［黒滝 紘生ら訳］『深層学習』ドワンゴ　2018.02
4. 金谷 健一『線形代数セミナー: 射影,特異値分解,一般逆行列』共立出版　2018.07

確率・統計

- 試行：1 回ずつあるいは 1 個ずつの個別の結果が偶然に左右される実験や観測のこと
- (根元)事象：試行によって起こりうる個々の結果のこと
  - 全事象（標本空間）：起こりうる全ての事象の集合
  - 和事象（和集合）：複数の事象のうち，少なくとも 1 つが起こるという事象
  - 積事象（積集合）：複数の事象が同時に起こるという事象
  - 空事象（空集合）：何も起こらないという事象
  - 余事象（補集合）：全事象の中である事象に含まれない事象
  - 互いに排反：任意の 2 つの事象が同時に起こり得ないこと
- 確率：事象の起こりやすさ
  1. 全事象の数と任意の事象の数との比率
  2. 試行回数と頻度から計算
  - 任意の事象 A に対し，　$0 \le P(A) \le 1$
  - 全事象　Ω　に対して，　$P(\Omega)=1$
  - 互いに排反な事象では，　$P(A \cup B)=P(A)+P(B)$
  - 互いに排反でない事象では，　$P(A \cup B)=P(A)+P(B)-P(A \cap B)$
- 条件付き確率
  排反でない事象 A, B で A が起こるという条件の下での B が起こる確率
  $$P(B|A)=\frac{P(A \cap B)}{P(A)}$$
- 独立：2 つの事象において，一方の事象が起こるかどうかが他方の事象の起こる確率に影響しない
  $$P(A \cap B)=P(A)P(B) \quad , \quad P(B|A)=P(B) \quad , \quad P(A|B)=P(A)$$
- ベイズの定理
  $$P(H_i|A)=\frac{P(H_i)P(A|H_i)}{\sum_{j=1}^{n} P(H_j)P(A|H_j)}$$
    - $P(H_i)$　：事前確率
    - $P(H_i|A)$　：事後確率
    - $P(A|H_i)$　：尤度
- 確率変数：変数の値が確率に基づき与えられる
- 確率分布：確率変数の値とその確率との対応関係
  - 離散型：とびとびの値ととる⇒ある値となる確率が存在する
  - 連続型：値が連続値⇒ある値となる確率は 0⇒ある範囲の値となる確率を求める
- 期待値：確率変数がどのような値をとることが「期待されるか」=確率変数の平均値
  - 離散型：　$E[X]=\sum_i x_i f(x_i)=\mu$
  - 連続型：　$E[X]=\int_{-\infty}^{\infty} x_i f(x_i)dx=\mu$
- 分散：確率分布のちらばり具合，確率変数の平均からの距離の 2 乗の期待値
  - 離散型：　$V[X]=E[(X-\mu)^2]=\sum_i (x_i-\mu)^2 f(x_i)=\sigma^2$
  - 連続型：　$V[X]=E[(X-\mu)^2]=\int_{-\infty}^{\infty} (x_i-\mu)^2 f(x_i)dx=\sigma^2$
- 標準偏差：分散は 2 乗しているので、平方根をとって単位を戻すことで比較が容易になる
- 共分散：2 つの確率変数の関係性を表す　$Cov[X,Y]=E[(X-\mu_x)(Y-\mu_y)]=\sigma_{xy}$
  - 2 つの確率変数に関係性が無いときは共分散は 0
- 共分散行列　$\Sigma=\begin{bmatrix} V[X] & Cov[X,Y] \\ Cov[Y,X] & V[Y] \end{bmatrix}$
- 代表的な確率分布
  - ベルヌーイ分布：2 種類の結果うち片方が確率 p で起こるときの確率分布
    $P(X=x)=p^x(1-p)^{1-x}$
  - 2 項分布：確率 p のベルヌーイ試行を n 回行ったとき，成功回数が x 回になる確率分布

$P(X=x)={}_nC_x\, p^x (1-p)^{n-x}$

- 正規分布：2 項分布の連続型への近似

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

- 推定量
  - 一致性：ある母数 θ の推定量 $\hat{\theta}$ が n を大きくすると θ となる
  - 不偏性：ある母数 θ の推定量 $\hat{\theta}$ の期待値が $E[\hat{\theta}]=\theta$ となる
- 標本平均の不偏推定量： $\bar{x}=\frac{1}{n}\sum x_i$　$E[\bar{x}]=\mu$　$V[\bar{x}]=\frac{\sigma^2}{n}$
- 分散の不偏推定量： $s^2=\frac{1}{n}\sum (x_i-\bar{x})^2$　$E[s^2]=\frac{n-1}{n}\sigma^2$
  - 不偏分散： $\hat{\sigma}^2=\frac{1}{n-1}\sum (x_i-\bar{x})^2$

【参考文献】

1. 日本統計学会編『改訂版 日本統計学会公式認定　統計検定 2 級対応　統計学基礎』東京図書　2015.12
2. Ian Goodfellow ら［黒滝 紘生ら訳］『深層学習』ドワンゴ　2018.02

情報理論

- 情報量：ある値を得た際の「おどろきの度合」
- 自己情報量：確率変数 X について，事象 X=x が起こったときの情報量 $I(x) = -\log P(x)$
- エントロピー：確率変数 X についての自己情報量の期待値

  $H(x) = E[I(x)] = \sum_x P(x)I(x) = -\sum_x P(x)\log P(x)$
- カルバック・ライブラーダイバージェンス：同じ確率変数 X に対して異なる確率分布$P(x)$と$Q(x)$がどれだけの差があるか

  $D_{KL}(P \parallel Q) = E_{x\sim P}[I(Q(x)) - I(P(x))] = \sum_x P(x)\log\frac{P(x)}{Q(x)}$
- クロスエントロピー

  $D_{KL}(P \parallel Q) = E_{x\sim P}[I(Q(x)) - I(P(x))] = E_{x\sim P}[I(Q(x))] - E_{x\sim P}[I(P(x))]$

  ここで，$H(P) = E_{x\sim P}[I(P(x))]$だから，$H(P,Q) = E_{x\sim P}[I(Q(x))]$とすると，

  $H(P,Q) = H(P) + D_{KL}(P \parallel Q)$

【参考文献】

1. Ian Goodfellow ら［黒滝 紘生ら訳］『深層学習』ドワンゴ　2018.02
2. C.M.ビショップ著[元田浩ら監訳]『パターン認識と機械学習 上』2012.01
3. 高村大也著『言語処理のための機械学習入門』コロナ社　2010.08

演習問題

1.1 $\boldsymbol{a} = \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix}$, $\boldsymbol{b} = \begin{pmatrix} 5 \\ 2 \\ 4 \end{pmatrix}$

問1.1.1 $\boldsymbol{a} + \boldsymbol{b} = \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix} + \begin{pmatrix} 5 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 6 \\ 8 \\ 7 \end{pmatrix}$

問1.1.2 $\boldsymbol{a} - \boldsymbol{b} = \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix} - \begin{pmatrix} 5 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} -4 \\ 4 \\ -1 \end{pmatrix}$

問1.1.3 $7\boldsymbol{a} = 7\begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix} = \begin{pmatrix} 7 \\ 42 \\ 21 \end{pmatrix}$

問1.1.4 $8(\boldsymbol{a} + \boldsymbol{b}) = 8\begin{pmatrix} 6 \\ 8 \\ 7 \end{pmatrix} = \begin{pmatrix} 48 \\ 64 \\ 56 \end{pmatrix}$

1.2 $A = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}$ $B = \begin{pmatrix} 1 & 4 \\ 1 & 5 \end{pmatrix}$

問1.2.1 $A + B = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 4 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 3 & 5 \\ 6 & 8 \end{pmatrix}$

問1.2.2 $A - 3B = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} - \begin{pmatrix} 3 & 12 \\ \cancel{15}3 & 15 \end{pmatrix} = \begin{pmatrix} -1 & -11 \\ 2 & -12 \end{pmatrix}$

2.1 $\boldsymbol{v} = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}$, $A = \begin{pmatrix} 1 & 3 & 4 \\ 5 & 9 & 0 \\ 3 & 1 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & 5 \end{pmatrix}$

問2.1.1 $A\boldsymbol{v} = \begin{pmatrix} 1 & 3 & 4 \\ 5 & 9 & 0 \\ 3 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 13 \\ 5 \\ 9 \end{pmatrix}$

問2.1.2 $B\boldsymbol{v} = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 10 \\ 15 \end{pmatrix}$

演習問題

問2.1.3 $BA=\begin{pmatrix}1&0&3\\0&2&5\end{pmatrix}\begin{pmatrix}1&3&4\\5&9&0\\3&1&2\end{pmatrix}=\begin{pmatrix}10&6&10\\25&23&10\end{pmatrix}$

問2.1.4 $B^T=\begin{pmatrix}1&0\\0&2\\3&5\end{pmatrix}$

2.2 $A=\begin{pmatrix}2&1\\4&1\end{pmatrix}$, $B=\begin{pmatrix}1&3\\3&1\end{pmatrix}$

問2.2.1 $AB=\begin{pmatrix}2&1\\4&1\end{pmatrix}\begin{pmatrix}1&3\\3&1\end{pmatrix}=\begin{pmatrix}5&7\\7&13\end{pmatrix}$

問2.2.2 $A^{-1}=\frac{1}{|A|}\begin{pmatrix}1&-1\\-4&2\end{pmatrix}=-\frac{1}{2}\begin{pmatrix}1&-1\\-4&2\end{pmatrix}$

$=\begin{pmatrix}-\frac{1}{2}&\frac{1}{2}\\2&-1\end{pmatrix}$

問2.2.3 $B^{-1}=\frac{1}{|B|}\begin{pmatrix}1&-3\\-3&1\end{pmatrix}=-\frac{1}{8}\begin{pmatrix}1&-3\\-3&1\end{pmatrix}$

$=\begin{pmatrix}-\frac{1}{8}&\frac{3}{8}\\\frac{3}{8}&-\frac{1}{8}\end{pmatrix}$

問2.2.4 $BAB^{-1}=\begin{pmatrix}1&3\\3&1\end{pmatrix}\begin{pmatrix}2&1\\4&1\end{pmatrix}\begin{pmatrix}-\frac{1}{8}&\frac{3}{8}\\\frac{3}{8}&-\frac{1}{8}\end{pmatrix}$

$=\frac{1}{8}\begin{pmatrix}-2&38\\2&26\end{pmatrix}$

演習問題

問 3.1　a, d

問 3.2

	0-4	1-3	2-2	3-1	4-0 ~~0-4~~
和	4	3	2	1	0
回数	75	300	450	300	75
確率	1/16	1/4	3/8	1/4	1/16

問 4.1　$I = -\log_2 P(x)$

問 4.1.1　$P(X=表) = \frac{1}{2}$　$I = -\log_2 \frac{1}{2} = \underline{1}$ //

問 4.1.2　$P(X=(表表)) = \frac{1}{4}$　$I = -\log_2 \frac{1}{4} = \underline{2}$ //

問 4.1.3　$P(X=1) = {}_nC_1 \left(\frac{1}{2}\right)^n = n\left(\frac{1}{2}\right)^n$

$I = -\log_2 n\left(\frac{1}{2}\right)^n = -\log_2 n - \log_2\left(\frac{1}{2}\right)^n$

$= \underline{n - \log_2 n}$ //

演習問題

問5.1 $\begin{cases} X: 洗濯物を干していた日 \\ Y: 雨が降ってきた日 \end{cases}$

$P(X) = \frac{60}{365}$, $P(Y) = \frac{7}{365}$

$P(X \cap Y) = \frac{12}{365}$

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} = \frac{\frac{12}{365}}{\frac{60}{365}} = \frac{12}{60} = \frac{1}{5} //$$

問5.2 $P(赤) = \frac{3}{5}$, $P(白) = \frac{2}{5}$ $P(A) = \frac{2}{5}$, $P(B) = \frac{2}{5}$

$P(C) = \frac{1}{5}$

問5.2.1 $$P(B|赤) = \frac{P(赤 \cap B)}{P(赤)} = \frac{\frac{1}{5}}{\frac{3}{5}} = \frac{1}{3} //$$

問5.2.2 $$P(白|A) = \frac{P(A \cap 白)}{P(A)} = \frac{\frac{1}{5}}{\frac{2}{5}} = \frac{1}{2} //$$

問6.1 $X = AB$, $\log X = \log AB = \underline{\log A + \log B} //$

問6.2 $X = \frac{A}{B}$ $\log X = \log \frac{A}{B} = \underline{\log A - \log B} //$

問6.3 $X = x_1 x_2 x_3 x_4$

$$\log X = \log x_1 x_2 x_3 x_4 = \log x_1 + \log x_2 + \log x_3 + \log x_4$$

$$= \underline{\sum_{k=1}^{4} \log x_k} //$$

演習問題

問7.1

$$\mathbb{a}=\begin{pmatrix}2\\6\\3\end{pmatrix},\ \mathbb{b}=\begin{pmatrix}1\\1\\4\end{pmatrix}\qquad \mathbb{a}+\mathbb{b}=\begin{pmatrix}2\\6\\3\end{pmatrix}+\begin{pmatrix}1\\1\\4\end{pmatrix}=\begin{pmatrix}3\\7\\7\end{pmatrix}$$

問7.2

$$A=\begin{pmatrix}1&4\\2&3\end{pmatrix}\qquad x=\begin{pmatrix}1\\1\end{pmatrix}$$

$Ax=\lambda x$ より $\begin{pmatrix}1&4\\2&3\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}5\\5\end{pmatrix}$

よって $\lambda x=\begin{pmatrix}5\\5\end{pmatrix}$ より $5\begin{pmatrix}1\\1\end{pmatrix}$ ∴ $\lambda=5$

問7.3 $\quad E(f)=\sum P(x)f(x)$

問7.4

$$V[f]=E\left[\left(f(x)-E[f(x)]\right)^2\right]$$

$$=E[f(x)^2]-\{E[f(x)]\}^2$$

問7.5 $\quad I(x)=-\log P(x)$

$$H(x)=\sum P(x)I(x)=-\sum P(x)\log P(x)$$