応用数学レポート

2024/5/15

1　線形代数

1.1　行列

行列とは、ベクトルを並べたもの。連立方程式の研究の中から生まれた。例えば、連立法定式 1 は、行列 A, ベクトル $\vec{b}$,$\vec{c}$ を用いて、式 2 のように $A\vec{x} = \vec{b}$ と整形できる。

$$\begin{cases} \ \ x_1 + 2x_2 = 3 \\ 2x_1 + 5x_2 = 5 \end{cases} \tag{1}$$

$$\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \tag{2}$$

また、行基本変形として、行列は以下の 3 つの変形ができる。

1. i 行目を c 倍する
2. s 行目に t 行目の c 倍を加える
3. p 行目と q 行目を入れ替える

1.2　単位行列と逆行列

単位行列 I とは、ある行列とかけても相手が変化しない行列のことであり、式 3 のように表される。すなわち、相手の行列を A とすれば、式 4 の関係になる。

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \tag{3}$$

$$AI = IA = A \tag{4}$$

行列 A の逆行列 A^{-1} とは、A とかけることで単位行列となる逆数のようなものであり、式 5 の関係になる。

$$AA^{-1} = A^{-1}A = I \tag{5}$$

逆行列の求め方は掃き出し法などがあり、逆行列が存在する条件として行列式が 0 でないということが挙げられる。

1.3 固有値と固有ベクトル

行列 A に対して、式 6 のようなベクトル $\vec{v}$ とスカラー値 λ があったとすると、$\vec{v}$,λ をそれぞれ行列 A に対する固有ベクトル、固有値という。

$$A\vec{v} = \lambda\vec{v} \tag{6}$$

1.4 固有値分解

正方行列 A に対し、固有値 $\lambda_1, \lambda_2, \cdots, \lambda_r$ と固有ベクトル $\vec{v_1}, \vec{v_2}, \cdots, \vec{v_r}$ を持ったとする。この固有値を対角線上に並べそれ以外の成分を 0 とした行列 Λ（式 7）と、固有ベクトルを並べた行列 V（式 8）を用意したとき、式 9 の関係となる。

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_r \end{pmatrix} \tag{7}$$

$$V = \begin{pmatrix} \vec{v_1} & \vec{v_2} & \cdots & \vec{v_r} \end{pmatrix} \tag{8}$$

$$AV = V\Lambda \tag{9}$$

すなわち、式 10 と変形できる。

$$A = V\Lambda V^{-1} \tag{10}$$

このように、正方行列 A を式 10 のように 3 つの行列の積に変換することを固有値分解という。固有値分解によって、行列の累乗の計算が容易になる等の利点があり、実際に行列 A の k 乗は式 11 と表せる。

$$A^k = V\Lambda^k V^{-1} = V \begin{pmatrix} {\lambda_1}^k & 0 & \cdots & 0 \\ 0 & {\lambda_2}^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {\lambda_r}^k \end{pmatrix} V^{-1} \tag{11}$$

1.4.1 問題（講義資料より一部変更して抜粋）

$\begin{pmatrix} 4 & 1 \\ 0 & 6 \end{pmatrix}$ を固有値分解せよ

$$A\vec{x} = \lambda\vec{x}$$
$$(A - \lambda I)\vec{x} = 0$$

ここで、$\vec{x} \neq \vec{0}$ から、$A-\lambda I$ は逆行列を持ってはいけない、すなわち、$|A-\lambda I| = 0$ であることから、

$$\begin{vmatrix} 4-\lambda & 1 \\ 0 & 6-\lambda \end{vmatrix} = 0$$
$$(4-\lambda)(6-\lambda) = 0$$
$$\lambda = 4, 6$$

$\lambda = 4$ のとき、

$$\begin{pmatrix} 4 & 1 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 4 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{よって、} x_2 = 0$$

$\lambda = 6$ のとき、

$$\begin{pmatrix} 4 & 1 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 6 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{よって、} 2x_1 = x_2$$

したがって、

$$\lambda = 4 \text{のとき} \vec{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{の定数倍}$$

$$\lambda = 6 \text{のとき} \vec{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{の定数倍}$$

つまり、

$$\begin{pmatrix} 4 & 1 \\ 0 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}^{-1}$$

$$\begin{pmatrix} 4 & 1 \\ 0 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} 1 & -1/2 \\ 0 & 1/2 \end{pmatrix}$$

1.5 特異値分解

正方行列以外で固有値分解と似たようなことができるように、特異値分解がある。以下、特異値分解について述べる。
正方行列でない行列 M に対して、式 12,13 を満たす単位ベクトル $\vec{u}, \vec{v}$ とスカラー値 σ があったとする。

$$M\vec{v} = \sigma\vec{u} \tag{12}$$

$$M^{\mathsf{T}}\vec{u} = \sigma\vec{v} \tag{13}$$

このとき、$\vec{u}$, $\vec{v}$ を行列 M に対する特異ベクトル、σ を特異値という。
さらに、それぞれ r 個の特異ベクトルと特異値をもっているとする、すなわち、特異値 $\sigma_1, \sigma_2, \cdots, \sigma_r$ と特異ベクトル $\vec{u_1}, \vec{u_2}, \cdots, \vec{v_r}, \vec{v_1}, \vec{v_2}, \cdots, \vec{v_r}$ を持ったとする。
特異値を対角線上に並べそれ以外の成分を 0 とした行列 S（式 14）と特異ベクトル $\vec{u}$,$\vec{v}$ をそれぞれ並べた行列 U（式 15）,V（式 16）を用意したとき、式 12 から式 17 の関係となる。

$$S = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix} \tag{14}$$

$$U = \begin{pmatrix} \vec{u_1} & \vec{u_2} & \cdots & \vec{u_r} \end{pmatrix} \tag{15}$$

$$V = \begin{pmatrix} \vec{v_1} & \vec{v_2} & \cdots & \vec{v_r} \end{pmatrix} \tag{16}$$

$$M = USV^{-1} \tag{17}$$

式 17 は、行列 M を固有値分解のように 3 つの行列の積に変換できており、この変換を特異値分解という。
一方、式 13 から、式 18 の関係も導かれる。

$$M^{\mathsf{T}} = VS^{\mathsf{T}}U^{-1} \tag{18}$$

そのため、式 17,18 より、式 19 が導かれる。

$$MM^{\mathsf{T}} = USV^{-1}VS^{\mathsf{T}}U^{-1} = USS^{\mathsf{T}}U^{-1} \tag{19}$$

すなわち、MM^{T} を固有値分解すれば、その左特異ベクトル U と特異値の 2 乗が求められることがわかる。同様に、$M^{\mathsf{T}}M$ を固有値分解すれば、その右特異ベクトル V と特異値の 2 乗を求められる。
特異値分解は、画像データ量を小さくするといった前処理で使われている。

1.5.1 問題（講義資料より抜粋）

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \text{を特異値分解せよ}$$

$$MM^{\mathsf{T}} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 14 & 10 \\ 10 & 14 \end{pmatrix}$$

これを固有値分解する。

$$\begin{aligned}
\begin{vmatrix} 14-\lambda & 10 \\ 10 & 14-\lambda \end{vmatrix} &= 0 \\
(14-\lambda)^2 - 10^2 &= 0 \\
(24-\lambda)(4-\lambda) &= 0 \\
\lambda &= 24, 4 \\
\lambda = 24 \text{ のとき} \begin{pmatrix} 14 & 10 \\ 10 & 14 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 24 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{ より} \vec{x} &= \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \\
\lambda = 4 \text{ のとき} \begin{pmatrix} 14 & 10 \\ 10 & 14 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 4 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{ より} \vec{x} &= \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \\
\text{よって} \begin{pmatrix} 14 & 10 \\ 10 & 14 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 24 & 0 \\ 0 & 4 \end{pmatrix} &\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}^{-1}
\end{aligned}$$

同様に、

$$M^{\mathsf{T}}M = \begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 8 & 6 \\ 8 & 8 & 8 \\ 6 & 8 & 10 \end{pmatrix}$$

を固有値分解する。

$$\begin{vmatrix} 10-\lambda & 8 & 6 \\ 8 & 8-\lambda & 8 \\ 6 & 8 & 10-\lambda \end{vmatrix} = 0$$

$$\lambda(24-\lambda)(4-\lambda) = 0$$

$$\lambda = 24, 4, 0$$

$$\lambda = 24 \text{ のとき} \begin{pmatrix} 10 & 8 & 6 \\ 8 & 8 & 8 \\ 6 & 8 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 24 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ より} \vec{x} = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}$$

$$\lambda = 4 \text{ のとき} \begin{pmatrix} 10 & 8 & 6 \\ 8 & 8 & 8 \\ 6 & 8 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 4 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ より} \vec{x} = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix}$$

$$\lambda = 0 \text{ のとき} \begin{pmatrix} 10 & 8 & 6 \\ 8 & 8 & 8 \\ 6 & 8 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ より} \vec{x} = \begin{pmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}$$

$$\text{よって} \begin{pmatrix} 10 & 8 & 6 \\ 8 & 8 & 8 \\ 6 & 8 & 10 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 24 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{pmatrix}^{-1}$$

つまり、

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 2\sqrt{6} & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \end{pmatrix}$$

と特異値分解できる。

2 確率・統計

2.1 条件付き確率

条件付き確率 $P(Y=y|X=x)$ とは、ある事象 $X=x$ が与えられた下で $Y=y$ となる確率である。$X=x$ と $Y=y$ が同時に発生する確率を同時確率 $P(Y=y, X=x)$ とすると、式 20 と表せる。つまり、同時確率を $X=x$ となる確率 $P(X=x)$ で割った値となる。

$$P(Y=y|X=x) = \frac{P(Y=y, X=x)}{P(X=x)} \tag{20}$$

2.2 ベイズ則

式 20 同様に、ある事象 $Y=y$ が与えられた下で $X=x$ となる確率 $P(X=x|Y=y)$ は 21 と表せる。

$$P(X=x|Y=y) = \frac{P(Y=y,X=x)}{P(Y=y)} \tag{21}$$

つまり、式 20,21 から式 22 が導かれる。

$$P(Y=y|X=x) = \frac{P(X=x|Y=y)P(Y=y)}{P(X=x)} \tag{22}$$

式 22 をベイズ則（ベイズの定理）という。
ベイズ則を利用することで、結果から原因を推定することができる。例えば、飴玉と子供たちの笑顔について以下の関係があるとする。

- ある街の子どもたちが飴玉をもらえる確率は 1/4
- 街の子どもたちが笑顔でいる確率は 1/3
- 飴玉をもらうと子供たちが笑顔になる確率は 1/2

ここで、笑顔な子どもがいたとしてその子が飴玉をもらっている確率は、ベイズ則から $1/2 \times 1/4 \div 1/3 = 3/8$ と求まる。よって、「飴玉をもらうと子供たちが笑顔になった」という結果から、笑顔の原因が飴玉である確率が 3/8 と求められ、約半分程度飴玉が理由であると推定できた。

2.3　期待値と分散

事象 $X = x_1, x_2, \cdots, x_n$ に対して、確率変数 $f(X) = f(x_1), f(x_2), \cdots, f(x_n)$、確率 $P(X) = P(x_1), P(x_2), \cdots, P(x_n)$ があったとする。このとき、確率変数 f に対する期待値 $E(f)$ は式 23 と表せる。すなわち、期待値とは確率変数の平均の値といえる。

$$E(f) = \sum_{k=1}^{n} P(X=x_k) f(X=x_k) \tag{23}$$

一方、データの散らばり具合を表す分散 $Var(f)$ は式 24 と表せる。

$$Var(f) = \sum_{k=1}^{n} P(X=x_k)(f(X=x_k) - E(f))^2 = E(f^2) - (E(f))^2 \tag{24}$$

さらに、2 つの事象系列 $X = x_1, x_2, \cdots, x_n$, $Y = y_1, y_2, \cdots, y_n$ に対する確率変数 f,g の傾向を表す共分散 $Cov(f,g)$ は式 25 と表せる。ここで、共分散が正の値を取れば 2 つの確率変数は似た傾向を、負の値を取れば逆（一方の確率が増加すれば、もう一方は減少する）の傾向にあるといえる。

$$Cov(f,g) = \sum_{k=1}^{n} P(X=x_k)(f(X=x_k) - E(f))(g(Y-y_k) - E(g)) = E(fg) - E(f)E(g) \tag{25}$$

2.4 様々な確率分布

2.4.1 ベルヌーイ分布

ベルヌーイ分布とは、確率 p で 1 を、確率 $q = 1 - p$ で 0 をとる確率分布のことである。つまり、結果が 2 つしかない試行の確率分布である。例えば、結果が表と裏しかないコイントスなどである。

2.4.2 マルチヌーイ分布

マルチヌーイ分布とは、ベルヌーイ分布を 2 つ以上の結果にまで拡張した確率分布である。例えば、6 面あるサイコロを回す試行などである。

2.4.3 二項分布

二項分布 $P(q|p, n)$ とは、試行回数 n と成功率 p に対して、成功回数 k の確率を表す確率分布である。ベルヌーイ分布の多試行版ともいえる。式 26 で表せる。

$$P(k|p, n) = {}_n\mathrm{C}_k p^k (1 - p)^{n-k} \tag{26}$$

2.4.4 ガウス分布（正規分布）

ガウス分布とは、平均を中心に左右対称の連続型の確率分布であり、正規分布とも呼ばれる。社会現象など様々な分野で現れることが多く、機械学習やパターン認識でも重要な役割を持つ。

3 情報理論

3.1 自己情報量

自己情報量とは、事象がもつ情報量（事象が起こる珍しさ）のことである。つまり、確率という言葉を用いれば以下の 2 つが成り立つ。

1. 確率が高いほど、自己情報量は小さくなる
2. 事象 x かつ y が起きた時の自己情報量は、事象 x と y の自己情報量の加算である

よって、事象 x の自己情報量 $I(x)$ として、対数 log を用いた式 27 が考えられる（底はなんでもよいが、以後ネイピア数とする）。

$$I(x) = -\log P(x) \tag{27}$$

3.2 シャノンエントロピー

シャノンエントロピー $H(X)$ とは、自己情報量の期待値である。式 28 で表せる。

$$H(X) = E(I(x)) = -\sum_{k=1}^{n} P(x_k) \log P(x_k) \tag{28}$$

3.3 KL ダイバージェンス

KL ダイバージェンス $D_{\mathbf{KL}}(P||Q)$ とは、同じ事象・確率変数における異なる確率分布 P,Q の違いを表す値であり、式 29 で表せる。

$$D_{\mathbf{KL}}(P||Q) = \sum_{k=1}^{n} P(x_k) \log \frac{P(x_k)}{Q(x_k)} \tag{29}$$

機械学習では真の確率分布 $P(x)$ が複雑な場合、よく知っている $Q(x)$ で近似して考えるアプローチが取られており、その際に「$Q(x)$ は $P(x)$ を近似できているか」を表す指標として KL ダイバージェンスが用いられている。

3.4 交差エントロピー

交差エントロピー $H(P,Q)$ とは、KL ダイバージェンス同様に 2 つの確率分布 P,Q の違いを表す値であり、式 30 で表せる。

$$H(P,Q) = -\sum_{k=1}^{n} P(x_k) \log Q(x_k) \tag{30}$$

交差エントロピーは、機械学習の損失関数などでよく用いられている。