応用数学　レポート

2024 年 06 月 06 日 新規作成

第一章 線形代数

1.1. スカラーとベクトル

- **スカラー**
  - 重さなどのように「大きさ」だけをもつ量
- **ベクトル**
  - 風などのように「向き」と「大きさ」をもつ量

1.2. 行列

次に述べるような数を箱に詰めたような形をしている。

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

- 例)

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix}, \begin{pmatrix} 1 & -2 & 3 \\ 2 & -4 & 6 \\ 3 & -6 & 9 \end{pmatrix}$$

- (m, n)型の行列（$m \times n$行列）

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

この行列Aに対して

$$\begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{pmatrix}$$

をAの第i行ベクトルといい、

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

をAの第j列ベクトルという。
(m, n)型の行列A, B, Cと実数λ, μについて以下が成り立つ。

(1) $A + B = B + A$ [交換法則]
(2) $A + (B + C) = (A + B) + C$ [結合法則]
$\lambda(\mu A) = \mu(\lambda A)$
(3) $\lambda(A + B) = \lambda A + \lambda B$ [分配法則]
$(\lambda + \mu)A = \lambda A + \mu A$ [分配法則]

また、(m, n)型の行列A、(n, p)型の行列B, Cと(p, q)型の行列Dおよび実数λについて以下が成り立つ。

(1) $A(BD) = (AB)D$　　　[結合法則]
(2) $A(B+C) = AB + AC$　　　[分配法則]
(3) $\lambda(AB) = (\lambda A)B = A(\lambda B)$

ここで、行列Aと行列Bの積ABは(Aの列の数) $=$ (Bの行の数)でないと定義できないので注意する。

1.3. 逆行列

　数字の 1 に相当する行列として右下がりの対角線上に 1 が並ぶ行列をn次元単位行列という。

$$E = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

実際、(m, n)型行列Aに対して$AE = A$が成り立つ。
n次元正方行列Rに対して、$RX = XR = E_n$を満たすn次元正方行列Xがあるとき、Rは正則であるといい、行列XをR^{-1}と表して、「Rインバース」と読み、Rの逆行列という。
2 次の正方行列Rを考える。

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$ad \neq bc$であれば、Rの逆行列R^{-1}は、

$$R^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

となる。
逆行列を使い、以下の連立方程式を解く。

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

これを行列で表すと

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

逆行列を左からかけ

$$\frac{1}{5-4}\begin{pmatrix}5 & -2\\ -2 & 1\end{pmatrix}\begin{pmatrix}1 & 2\\ 2 & 5\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \frac{1}{5-4}\begin{pmatrix}5 & -2\\ -2 & 1\end{pmatrix}\begin{pmatrix}3\\ 5\end{pmatrix}$$

よって

$$\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}5\\ -1\end{pmatrix}$$

1.4. 行列式

行列式（Determinant）とは、行列がどのような特性を持っているかを示す数値のこと。2 次の行列式の絶対値はその成分が作るベクトルによって構成される平行四辺形の面積に等しい。3 次の場合は平行六面体の体積に一致する。行列式はスカラーになる。
2 次の正方行列の場合、

$$A = \begin{pmatrix}a & b\\ c & d\end{pmatrix}$$

Aの行列式$\left|A\right|$は以下のように計算する。

$$\left|A\right| = ad - bc$$

3 次の正方行列Bの場合、行列式$\left|B\right|$は以下のように計算する。

$$\begin{aligned}\left|B\right| &= \begin{vmatrix}b_{11} & b_{12} & b_{13}\\ b_{21} & b_{22} & b_{23}\\ b_{31} & b_{32} & b_{33}\end{vmatrix}\\ &= b_{13}b_{21}b_{32} + b_{11}b_{22}b_{33} + b_{12}b_{23}b_{31}\\ &\quad -b_{12}b_{21}b_{33} - b_{13}b_{22}b_{31} - b_{11}b_{23}b_{32}\end{aligned}$$

このような計算方法をサラスの方法という。これは 4 次以上になると計算が困難になるため 3 次までしか使えない。また、別解として

$$\begin{aligned}\left|B\right| &= \begin{vmatrix}b_{11} & b_{12} & b_{13}\\ b_{21} & b_{22} & b_{23}\\ b_{31} & b_{32} & b_{33}\end{vmatrix}\\ &= b_{11}\begin{vmatrix}b_{22} & b_{23}\\ b_{32} & b_{33}\end{vmatrix} + b_{12}\begin{vmatrix}b_{21} & b_{23}\\ b_{31} & b_{33}\end{vmatrix} + b_{13}\begin{vmatrix}b_{21} & b_{22}\\ b_{31} & b_{32}\end{vmatrix}\end{aligned}$$

が成り立つ。このように行列式の次数を下げることを余因子展開という。

1.5. 行基本変形

以下に行列に関する 3 つの操作を示す。この 3 つの操作を行基本変形という。

- (I) 第i行を$\lambda(\neq 0)$倍する
- (II) 第i行に第k行を加える
- (III) 第i行と第k行とを入れ替える

また、(I)と(II)を組み合わせた「第i行に第k行のλ倍を加える」こともできる。
行基本変形により逆行列を求める。

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$$

$$\begin{aligned} A^{-1} &= \begin{pmatrix} 1 & 2|1 & 0 \\ 2 & 5|0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2|1 & 0 \\ 0 & 1|-2 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0|5 & -2 \\ 0 & 1|-2 & 1 \end{pmatrix} \end{aligned}$$

計算の 1 行目->2 行目では行列の 1 行目を-2 倍したものを 2 行目に足している。
次に計算の 2 行目->3 行目では行列の 2 行目を-2 倍したものを 1 行目に足している。
ここで右側の2×2の行列が逆行列となっていることがわかる。同様にして連立方程式も解ける。このような手法を掃き出し法という。

1.6. 固有値分解

n次正方行列Aに対して、

$$Av = \lambda v, \quad v \neq 0$$

を満たすとき、vは方向を変えないベクトルとなりこれを固有ベクトルといい、拡大・縮小率を表すスカラーλを固有値という。この時、

$$(A - \lambda E)v = \mathbf{0}$$

は$\mathbf{0}$以外の解を持つ。係数行列の行列式は$\left|A - \lambda E\right| = 0$である。ここである正方行列$P$を用いて、

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix} = \Lambda$$

と対角行列にできるとき、Aは対角可能であるといい、

$$A = P\Lambda P^{-1}$$

となるようにAを分解することを固有値分解という。Pは固有ベクトルをの組を列ベクトルとする行列になる。固有値分解は主成分分析やサポートベクターマシンで使われる。また、行列のべき乗や指数関数などの計算が容易になることがメリットとして挙げられる。

例）

$$A = P\Lambda P^{-1} \quad \Rightarrow \quad A^n = P\Lambda^n P^{-1}$$

1.7. 特異値分解

　固有値分解できるのはAが対角可能である場合に限られたが、特異値分解は任意の$(n \times m)$行列に対して適用できる。特異値分解では行列Aを以下のように分解する。

$$A = U\Sigma V^T$$

ここで

- Uは$(n \times n)$の直行行列で、その列はAA^Tの固有ベクトルである。
- Σは$(n \times m)$の対角行列で対角要素はAの特異値となる。また、特異値はAの特異値分解から得られる固有値の平行根となる。
- Vは$(m \times n)$の直行行列でその列はA^TAの固有ベクトルである。

特異値分解の手順

1. 行列AA^TとA^TAを計算する。
   $AA^T = U\Lambda U^T$
   $A^TA = V\Lambda V^T$
   ここでΛは対角行列でその対角要素はAの特異値の 2 乗である。
2. 固有値と固有ベクトルの計算
   UはAA^Tの固有ベクトル行列
   VはA^TAの固有ベクトル行列
   これらの固有ベクトルがそれぞれ左特異ベクトル、右特異ベクトルである。
3. 特異値はΛの対角要素の平方根となる。
   $\sigma_i = \sqrt{\lambda_i}$
4. 対角行列Σの形成
   対角行列Σは特異値を対角要素に持つ。

第二章 確率・統計

2.1. 確率

　確率とは事象の起こりやすさを定量的に示すものであり、事象Aの確率を probability の頭文字をとって$P(A)$で表す。試行が全部でN個あって、同程度の確かさで起こる事象AがR回起こったとすれば、確率は$P(A)=R/N$と定義される。これは客観的に決定されるので、客観説の立場という。これに対して、主観的にある確率を与えて主観確率を分析することを主観説の立場という。

2.2. 条件付確率

　事象Aの起こる確率が他の事象Bに影響されない場合、以下が成り立つ。

$$P(A \cap B) = P(A) \cdot P(B)$$

事象Bが起こったとわかっている場合に、事象Aの起こる確率をBを条件とするAの条件付確率と呼び、$P(A|B)$で表す。Aの条件付確率は、Bが起こった場合にそのうちさらにAが起こる確率であるから、

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

と定義される。同様に

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

となることから、$P(A \cap B)$を消去すると、

$$P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)}$$

が得られる。これをベイズの定理という。ベイズの定理は結果に対する原因の確率を計算する公式を与える。

2.3. 確率変数と確率分布

- 確率変数
  ランダムにある値をとる変数。Xのように慣例的に大文字で表す。
- 確率分布
  確率変数がとりうる値それぞれの確率を表したもの。

2.4. 確率変数の期待値と分散

いろいろな値を取る確率変数を代表する確率の重みつき平均を期待値$E(X)$という。

- 離散型

$$E(X) = \sum_x xf(x)$$

- 連続型

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

確率分布の集中やばらつきを示す指標を分散という。分散$V(X)$は期待値μを使い以下のように定義される。

$$V(X) = E\{(X-\mu)^2\}$$

- 離散型

$$V(X) = \sum_x (x-\mu)^2 f(x)$$

- 連続型

$$V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

また、分散の定義から

$$\begin{aligned} V(X) &= E\{(X-\mu)^2\} \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

となり、二乗期待値と期待値の二乗から計算できる。また、標準偏差$D(X)$は分散の平方根で表される。

$$D(X) = \sqrt{V(X)}$$

2.5. 共分散

二つの確率変数X, Yの間に関連がある場合を考えると、ばらつきの指標としての分散には単純な加法が成立しない。

$$V(X+Y) \neq V(X)+V(Y)$$

$X+Y$のばらつきにはX, Y単独のばらつきのほかに相互関連によるばらつき$\mathrm{Cov}(X, Y)$が存在する。

$$\begin{gathered}V(X+Y)=V(X)+V(Y)+2\mathrm{Cov}(X, Y) \\ \mathrm{Cov}(X, Y)=E\{(X-E(X))(Y-E(Y))\}\end{gathered}$$

$\mathrm{Cov}(X, Y)$をX, Yの共分散という。共分散をX, Yの標準偏差で割って調整した値を相関係数という。

$$\rho_{XY}=\frac{\mathrm{Cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

2.6. 確率分布の例

- 二項分布
  成功するか失敗するか、のような 2 種類の可能な結果が生じる事象をn回繰り返すことをベルヌーイ試行という。成功確率p、失敗確率$1-p$とすると、確率分布は以下のようになる。

  $$f(x)={}_n C_x p^x(1-p)^{n-x}$$

  $$\begin{aligned}E(X)&=np \\ V(X)&=np(1-p)\end{aligned}$$

  $n=1$の時の$f(x)$をベルヌーイ分布という。
- ポアソン分布
  二項分布においてnが大である一方、pが希少現象である場合にはポアソン分布となる。

  $$f(x)=e^{-\lambda}\frac{\lambda^x}{x!}$$

  $$\begin{aligned}E(X)&=\lambda \\ V(X)&=\lambda\end{aligned}$$

  これは二項分布において$np \rightarrow \lambda$となるように$n \rightarrow \infty, p \rightarrow 0$となる極限を求めると得られる。

- 正規分布
  連続型の代表的な確率分布でガウス分布ともいう。中心極限定理と関連することから統計学において重要な分布である。

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

$$E(X) = \mu$$
$$V(X) = \sigma^2$$

第三章 情報理論

3.1. 自己情報量

特定の出来事が発生する際の驚きの度合いを定量化する指標。ある出来事xが発生する確率を$P(x)$とすると、自己情報量$I(x)$は次のように定義される。

$$I(x) = -\log_2 P(x)$$

対数の底が 2 の時、単位は bit となり、ネイピア数のときは nat となる。出来事の確率が低いほど自己情報量は大きくなる。これは稀な出来事が発生するほど驚きが大きいことを反映している。

3.2. シャノンエントロピー

確率分布全体の不確実性や情報量を表す。ある離散型確率分布Pのシャノンエントロピー$H(P)$は次のように定義される。

$$H(P) = -\sum_i P(x_i)\log P(x_i)$$

この値が高いほど不確実性や情報量が大きいことを意味する。

3.3. カルバックライブラーダイバージェンス

ある確率分布Pから別の確率分布Qへの情報のずれや差異を測定する指標。PとQの KL ダイバージェンス$D_{KL}(P \mid\mid Q)$は次のように定義される。

$$D_{KL}(P \mid\mid Q) = \sum_i P(x_i)\log\frac{P(x_i)}{Q(x_i)}$$

KL ダイバージェンスは非対称であり、$D_{KL}(P \parallel Q) \neq D_{KL}(Q \parallel P)$である。KL ダイバージェンスは分布$P$に基づく情報を分布$Q$で近似する際の情報損失を示す。$P$と$Q$が同じであれば KL ダイバージェンスは 0 になる。

3.4. 交差エントロピー

　ある確率分布Pを使って別の確率分布Qを符号化する際の平均的な情報量を示す。PとQの交差エントロピー$H(P,Q)$は次のように定義される。

$$H(P,Q) = -\sum_i P(x_i) \log Q(x_i)$$

交差エントロピーは分布Qを用いて分布Pを表現する際の効率を示す。QがPに近いほど交差エントロピーは小さくなる。