第1章：線形代数

ベクトル

大きさと向きと持ちスカラーとのセットで表示される。

行列

数や記号や式などを縦と横に並べたもの。

行列の積

行列Aと行列Bの積については、$A \cdot B \neq B \cdot A$であり可換ではない。

単位行列

掛けても掛けられても相手が変化しない正方行列のこと。

単位行列はIで表す。

対角成分に1が並び、他の成分は全て0となるので、行列要素をa_{ij}とすると各要素は以下のように表すことができる。

$$a_{ij} = \begin{cases} 1 \ (\ i = j) \\ 0 \ (\ i \neq j) \end{cases}$$

例) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

逆行列

行列Aと掛けても掛けられても単位行列となる行列を行列Aに対する逆行列という。

逆行列はA^{-1}と表記する。

$$AA^{-1} = A^{-1}A = I$$

$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$の逆行列$A^{-1}$は以下の式で求めることができる。

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

上記より、$ad - bc = 0$だと逆行列は存在しない。

<掃き出し法による逆行列の求め方>

1. n行n列の行列Aに対してn行n列の単位行列Iの行列を合体させた$(A|I)$という行列を作成する。
2. $(A|I)$という行列に行基本変形を行い、行列の左半分が単位行列Iとなるように変形をする。
3. 行列の左半分が単位行列となった時の右半分の行列を抜き出すと逆行列A^{-1}となっている。

行列式

逆行列が存在するかしないかを判別するための式。

|A|や det(A)と表記する。

$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$については、$|A| = a_{11}a_{22} - a_{12}a_{21}$

固有値と固有ベクトル

A：正方行列、λ：スカラ、$\vec{x}$：0 ベクトルでないベクトルで以下が成り立つとき、

$A\vec{x} = \lambda\vec{x}$

λを行列Aの固有値、$\vec{x}$を行列Aの固有ベクトルと言う。

例) $\begin{pmatrix} 2 & 6 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 8\begin{pmatrix} 1 \\ 1 \end{pmatrix}$　固有値：8、固有ベクトル$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

固有値と固有ベクトルの求め方

$A\vec{x} = \lambda\vec{x}$を変形して、$(A - \lambda I)\vec{x} = \vec{0}$

$\vec{x} \neq \vec{0}$より、$|A - \lambda I| = 0$となり、これを満たす固有値：λを求め、その固有値に対する固有ベクトル：$\vec{x}$を求める。

固有値分解

A：正方行列、λ_i：行列 A の固有値、$\vec{v_i}$：行列 A の固有値λ_iに対する固有ベクトルに対して、以下の変形をすること。

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

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \ddots \end{pmatrix}$$

$$V = (\overrightarrow{v_1} \quad \overrightarrow{v_2} \quad \cdots)$$

固有値分解すると行列の累乗の計算が容易になる。

特異値分解

正方行列ではない行列に対して、固有値分解に似たような分解をすること。

行列が正方行列でない場合でも、特異値分解することで固有値分解のメリットである行列の累乗の計算が容易になる。

$MV = US$

$M^T U = VS^T$

$M = USV^{-1}$

$M^T = VS^T U^{-1}$

$MM^T = USV^{-1}VS^T U^{-1} = USS^T U^{-1}$

M：m行n列の非正方行列、U：m行m列のユニタリ行列、S：m行n列の非正方行列、V^{-1}：n行m列の非正方行列Vの逆行列

<関連記事レポート>

第 2 章：確率・統計

頻度確率(客観確率)

観測された頻度分布や想定された母集団の割合から導かれる確率のこと。

ベイズ確率(主観確率)

確率の概念を解釈したもので、ある現象の頻度や傾向の代わりに、確率を知識の状態を表す合理的な期待値、個人的な信念の定量化した解釈としたもの。

真偽が不明な命題を用いた推論を可能にするために用いる。

条件付き確率

ある事象 A が起こるという条件のもとで、別の事象 B が起こる確率のこと。

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

P(Y|X)　　　：事象 *X* が起きた後での事象 *Y* の確率。

P(Y, X)　　　：事象 *X* と事象 Y が同時に起こる確率。

P(X)　　　　：事象 *X* が起きる確率。

独立な事象の同時確率

お互いの発生確率に因果関係がない事象について、同時に発生する確率のこと。

$$P(X=x,Y=y)=P(X=x)P(Y=y)=P(Y=y,X=x)$$

P(X)：事象 *X* が起きる確率。

因果関係がないので、それぞれの事象が発生する確率の積が同時に発生する確率となる。

ベイズ則

関連性のある 2 つの事象 A, B について、事象 A が起こる前に事象 B が起こる確率である P(B)、事象 B が起こる前に事象 A が起こる確率である P(A)、事象 A が起きてから事象 B が起きる確率である P(B|A)、事象 B が起きてから事象 A が起きる確率である P(A|B)とした場合に、これらの関連性を表す定理。

$$P(X=x|Y=y)P(Y=y)=P(Y=y|X=x)P(X=x)$$

P(X)：事象 *Y* が起きる前の事象 *X* の確率(事前確率)。

P(X/Y)：事象 *Y* が起きた後での事象 *X* の確率(事後確率(条件付き確率))。

期待値

確率変数の全ての値に確率の重みを付けた加重平均。確率分布に対して定義する場合は、平均と呼ばれることが多い。

確率変数とは各事象に結び付けられた数値のこと。(例：サイコロを振った時に出る目の値など)

確率分布は確率変数に対する分布のこと。

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

$E(f)$：期待値、$P(X)$：確率、$f(X)$：確率変数

分散

データの散らばり具合を表す。

値が大きいほど散らばり具合が大きい。0 の場合はデータの値が全て等しいことと同じ。

標準偏差を 2 乗したものであり、計算が簡単となるため利用することが多い。

$$Var(f) = E\left(\left(f_{(X=x)} - E_{(f)}\right)^2\right) = E\left(f_{(X=x)}^2\right) - \left(E_{(f)}\right)^2$$

$E\left(f_{(X=x)}^2\right)$：$x$の 2 乗の平均値、$\left(E_{(f)}\right)^2$：$x$の平均値

共分散

2 種類のデータの相関関係を表す。

共分散の値が正の場合：2 種類のデータは正の相関(一方が増加するともう一方も増加する)。

共分散の値が負の場合：2 種類のデータは負の相関(一方が増加するともう一方も減少する)。

共分散の値が 0 の場合：2 種類のデータは相関がない(関係性が乏しい)。

$$Cov(f,g) = E\left(\left(f_{(X=x)} - E_{(f)}\right)\left(g_{(Y=y)} - E_{(g)}\right)\right) = E(fg) - E(f)E(g)$$

標準偏差

平均値からの散らばり具合を表す指標の一つ。

分散の平方根であり、分散は 2 乗しているため元データと単位が異なるが、標準偏差は元データと同じ単位となる。

$$\sigma = \sqrt{Var(f)} = \sqrt{E\left(\left(f_{(X=x)} - E_{(f)}\right)^2\right)}$$

ベルヌーイ分布

成功・失敗、表・裏のように 2 種類のみの結果しか得られない試行(ベルヌーイ試行)の結果を 0 と 1 で表した分布。

$P(x \mid \mu) = \mu^x (1 - \mu)^{1-x}$

$P(x \mid \mu)$：確率質量関数、μ：平均、x：パラメータ(0 or 1)

マルチヌーイ(カテゴリカル)分布

ベルヌーイのように 2 種類のみの結果ではなく、3 種類以上の結果が得られる試行の確率分布。

二項分布

ベルヌーイ試行を独立に複数介した時の分布。

$$P(x;\ \lambda, n) = \frac{n!}{x!\,(n-x)!}\,\lambda^{x}(1-\ \lambda)^{n-x}$$

$P(x;\ \lambda, n)$：確率質量関数、λ：事象の起こる確率、n：試行回数、x：事象の起こる回数

n 回試行した時に、事象が x 回起こる確率を表したもの。

ガウス分布(正規分布)

平均値と最頻値と中央値が一致する左右対称な分布。

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

$N(x;\ \mu, \sigma^2)$：確率密度関数、μ：平均、σ^2：分散

3 章：情報理論

自己情報量

事象そのものの情報量のこと。

$I(x) = -\log\big(P(x)\big) = \log(W(x))$

$I(x)$：事象xの自己情報量

$P(x)$：事象xの確率

$W(x)$：$\frac{1}{P(x)}$

確率の逆数の対数であるため、確率が小さい事象ほど自己情報量は大きい。

シャノンエントロピー(=微分エントロピー)

様々な事象の情報量の平均のこと。

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

$H(x)$：事象xの情報量

$E(I(x))$：$I(x)$の期待値

$I(x)$：事象xの自己情報量

$P(x)$：事象xの確率

カルバック・ライブラー　ダイバージェンス

2 つの確率分布 P, Q の差異を表す尺度。

$$D_{KL}(P\|Q) = \sum_x P(x)(-\log\left(Q(x)\right)) - \left(-\log\left(P(x)\right)\right) = \sum_x P(x) log \frac{P(x)}{Q(x)}$$

交差エントロピー

確率分布：P、別の確率分布：Qとした場合に、Q についての自己情報量を Pの分布で平均化したもの。

2 つの確率分布が似ている場合は、交差エントロピーの値は小さくなる。

クラス分類ニューラルネットワークの損失関数として利用される。

$H(P, Q)$　　：Qの Pに対する交差エントロピー

$H(P)$　　　　：Pのエントロピー

$D_{KL}(P//Q)$　：P から Q のカルバック・ライブラー ダイバージェンス

$H(P,Q) = H(P) + D_{KL}(P\|Q) = -\sum_x P(x)\log Q(x)$