

Report 1-3

第1章：線形代数

線形代数とは？

線形代数は、ベクトル空間と線形変換を中心とした学問体系

膨大なデータや複雑なシステムを簡明に表現でき

コンピュータで計算しやすいため発展した

分野を問わず広く応用されている。

■行列

数 → スカラー

スカラーとスカラーの組み合わせ → ベクトル

ベクトルとベクトルの組み合わせ → 行列

行列と行列の組み合わせ → テンソル *アインシュタインが一般相対性理論発表のために開発

行(横)方向

列(縦)方向

行列のそれぞれの方向は、漢字の「つくり」に出てくる二本線の方角で覚えると楽

行に同じ単位のを並べ、

列で種類を分けることが多い＝レコード？

*表計算やデータベースなどで(例：名簿)の場合は反対で、横にレコードを配置し、列(縦)に各種要素の単位であるフィールドを並べている

レモンティー（ホット及びアイス）の価格とカロリー

Lサイズやハーフサイズなどを設定すると、一度に計算できる

- 行列の四則演算

足し算、引き算、掛け算はあるが割り算はない

掛け算に関しては不可換

割り算は単位行列と逆行列を使う

- 逆行列

掛けると1となるような行列

割り算の代わりに利用

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}$$

2×2の逆行列の求め方

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

(ad-bc)=0 の時、逆行列は存在しない

(ad-bc)のことを行列式と呼ぶ

3x3の時はフオーク方式

aei+bfg+chd -ceg-bdi-ahf 前半足し算 後半引き算

逆行列の求め方 掃き出し法

連立方程式は

行列で表現できる

(ad-bc)=0 の時

不定 解がたくさんある

不能 解けない

などになる

連立方程式を
行列を使って解くメリット

連立方程式を手計算で解く際はいちいち考えて
計算しないといけないが、行列を使って解く場合は
公式に当てはめるだけなのでコンピュータ処理などに強い

- ・一次変換

点を別の点に変換、というよりも
空間を変換する

複数の点を一次変換すると

回転や拡大縮小が出来る。

＊テレビゲームなどで利用される

【参考】

マンガでわかる機械学習

マンガ線形代数入門

データサイエンティスト検定リテラシーレベル対策

■固有値分解

主成分分析による次元削減で利用

ケーリーハミルトンの定理

$$AA^T - (a+b)A + (ad-bc)E = 0$$

λ, \vec{x} を行列 A の固有値、固有ベクトルとする。

固有ベクトルは、変換後も方向が変わらない

λ を大きい順で並べ替え、これを対角にする対角ベクトル Λ を作る。

さらに、対応する \vec{x} を同順で列にして行列 V をつくと

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

なる行列 A をとることができる。これを行列 A の固有値分解という

■特異値分解

次元削減で利用

正方行列以外は固有値分解できないため、

その拡張として正方行列以外の行列には特異値分解があり固有値分解のメリットを享受できる。

m行n列の行列**M**は

$$\mathbf{M} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}$$

の形に分解する事ができる。

U : 行列Mの入力の基底となる正規直交ベクトル (m次ユニタリ行列)

V : 行列Mの出力の基底となる正規直交ベクトル (n次ユニタリ行列)

Σ : 行列Mの特異値を対角成分にもつm行n列行列。

第2章：確率・統計

分散 と 偏差

相関係数

「相関関係」があることと「因果関係」があること は一緒ではない

「相関関係がある（アイスクリームの販売数と溺死事故)けど、因果関係がない」

- $1 \leq \text{相関係数} r \leq 1$

$0.7 \leq |r| \leq 1$ 強い相関

$0.4 \leq |r| < 0.7$ 相関あり

$0.2 \leq |r| < 0.4$ 弱い相関

$0 \leq |r| < 0.2$ ほとんど相関がない

「相関係数 = 0 → 変数間に関係がない」とは言えない

例

変数をプロットしていくと円になる

「相関係数 = 0 → 変数間に線形関係はない」

「相関係数 = 1 → 因果関係がある」とは言えない

ポストの数とインフルエンザの患者数

ポストの数を減らしてもインフルエンザの患者数は減らない

確率変数の型 (確率分布が変わってくる)

離散型 とびとびの値 例：コイン投げの裏表、来店者数 → 確率質量

確率分布：ベルヌーイ分布、二項分布、ポアソン分布

連続型 連続した値 例：身長、体重 → 確率密度

確率分布：一様分布、正規分布、標準正規分布、指数分布、カイ二乗分布

統計学の種類

記述統計学 データを整理して、その特徴を明らかにする

推測統計学 標本データから母集団の特徴を推定・仮説検定する

【参考】

マンガでわかる統計学入門

データサイエンティスト検定リテラシーレベル対策

第3章：情報理論

クロード・シャノン(1916-2001) 情報理論の考案者

コンピューターの3大偉人(ノイマン、チューリング)の中で最も知名度が低い

これまで曖昧だった情報を定量的に扱えるようにした

自己情報量: $i(x) = -\log_2 P(x)$

対数の底が2の時、単位はbit * コンピューターで扱うのはこっち

対数の底がネイピア数eの時、単位はnat

例：

「わたし、パクチー大好き！」 パクチー好きが約1/4だった場合の情報量
 $-\log_2 1/4 = 2$ 情報量は2bit

「わたし、結婚詐欺にあった」 結婚詐欺にあった人が1/1024場合の情報量
 $-\log_2 1/1024 = 10$ 情報量は10bit

「わたし、美味しいものが好き」 美味しいものが好きな人は全員
 $-\log_2 1/ = 0$ 情報量は0bit

データ

質的変数 数値でないデータ

名義尺度 例：名前、性別、国籍

順序尺度 例：成績（優・良・可・不可）

量的変数 数値データ

間隔尺度 例：気温、西暦

比尺度 原点(0)が意味を持つ、何倍という概念が存在

例：時間、身長、音量、売上、金額、個数

参考)

データサイエンティスト検定リテラシーレベル対策

明日使えるインテリ悪口

