

# 応用数学

## 線形代数

### ベクトル

- スカラーのセット。大きさと向きを持つ。

### 行列

#### 行列とはなんなのか

- スカラーを表にしたものあるいはベクトルを並べたもの
- ベクトルの演算つまり変換に使用することができる。
  - バリエーションを作るために行列式？
  - 連立方程式の例で説明。
    - 行列を使うことで（係数を）シンプルに説明
- どのように使うか。
- 行列とベクトルの積
  - 新しい成分は古い成分全てから影響を受ける。故に変換することができる考える。
  - 行列と行列の積
    - 行列がベクトルの集まりと考えれば同様に計算できる。
    - 行と列を掛け合わせて新たな成分を作ると考えることができる。
- 連立方程式の計算と行列の計算の関係
  - 数式で連立方程式を解く手順は行列（行基本変形）で表すことができる。
  - 行列は左からかける事で表せる。
  - かける行列式には法則がある。
    1. 単位行列の*i*行目を*c*倍した行列をかけることで、対象の任意の行目を*n*倍できる。
    2. *s*行目に*t*行目の*c*倍を加える場合、単位行列の(*s*,*j*)成分を*c*倍する。
    3. *p*行目と*q*行目を入れ替える場合は*p*,*q*行目の対角成分を0とし、代わりに(*p*,*g*)及び(*q*,*p*)成分を1とする。
- 行列の「逆数」のようなもの

- 上記1.-3.について必要なものを左から順にかけていた物を計算してまとめることで、答えを一度に導き出せる行列が表せられる。
- 逆行列という。スカラー計算における割り算のように使える。
- 逆行列は  $ax=b$  を  $x= b/a$  とできるようにする役割を持つ。
- 単位行列
  - 対角成分が1で他が0の行列
  - 掛け算に使用しても元の行列の性質を変化させない
  - $A \cdot A^{-1}$  の-1はインバースと呼び、元の行列を打ち消す行列のことを言う。
- 逆行列の求め方
  - 単位行列を使って、左右の式の形を合わせる（式変形をわかりやすくできる）
  - 拡大行列と呼ばれるもので行基本変形を記録できる
  - 掃き出し法は拡大行列の行基本変形を繰り返すことで逆行列を求める手法。
- 逆行列が存在しない場合
  - 数字で逆数が存在しないケース（0）と同様に考える
  - 例えば、1つの解に定まらない連立方程式の行列表現
  - 2行2列で考える場合、 $a:b = c:d$  つまり  $ad - bc = 0$  の時は逆行列が存在しない。
  - あるいはそれぞれのベクトルで表される平行四辺形の面積が0の時
- 逆行列を持つか持たないか判別する式（determinant）→行列式
  - 行列式一般の性質
    - 同じ行ベクトルが含まれていると行列式はゼロ
    - 1つのベクトルが $\lambda$ 倍されると行列式は $\lambda$ 倍される
    - 他の成分が全部同じでi番目のベクトルだけが違った場合、行列式の値は元の行列式の値とi番目のベクトルが元の行列式との差分になっているものの行列式（他の成分は同一）の和になる。
    - 行列式のi成分とj成分のベクトルを入れ替えると符号が入れ替わる（上記3つの性質から導出できる）→ 証明添付（方法はスクショ）
  - 3つ以上のベクトルができるている行列式は展開できる
  - 行列式は以下。

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = \begin{vmatrix} a & b & c \\ 0 & e & f \\ 0 & h & i \end{vmatrix} + \begin{vmatrix} 0 & b & c \\ d & e & f \\ 0 & h & i \end{vmatrix} + \begin{vmatrix} 0 & b & c \\ 0 & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - d \begin{vmatrix} b & c \\ h & i \end{vmatrix} + g \begin{vmatrix} b & c \\ h & i \end{vmatrix}$$

- 行列式の解き方
  - 3次元は2次元に展開して解ける。
  - どの縦の列で2次元に展開してもよい。
  - また、転置しても行列式の値は変わらないので、横の列で展開しても答えは同じ。

- どこを展開すると計算式が少なくなるのかを考える。

## 固有値・固有値分解・特異値・特異値分解

### • 固有値

- ある行列Aに対して特殊なベクトルvをかけると元のベクトルvの $\lambda$ 倍になってしまうv,  $\lambda$ が存在することがある。このようなx,  $\lambda$ をそれぞれ、固有ベクトル、固有値と呼ぶ。
- 固有ベクトルはある特定の値というよりはある特定の比率になるので、1つに決まらない。固有ベクトルは「ベクトル〇〇の定数倍」などと表現したりする。一方固有ベクトルは一意に定まる。

### • 固有値分解

- 分解することで分類することができる。
  - 似たような特徴を見つけ出すことができる。
- 固有値は次元の数だけ存在することが知られている。
- とある行列Aに対して、固有値を体格に並べた行列 $\Lambda$ を用意して、各固有値に対応する固有ベクトルを横に並べた行列Vを考える。
  - $\Lambda$ の対角成分は各 $\lambda$ を大きい順に並べると良い。(わかりやすい)
- $AV$ は $A$ vを横に並べたものと考えることができる。
- 同様に $\lambda v$ を表すものは $V\Lambda$
- $AV = V\Lambda$ の関係性が成り立つ
- 変形して、 $A = V\Lambda V^{-1}$
- ベクトルを固有値と固有ベクトルの関係に分解できる。これを固有値分解という。
- 固有値分解を行うことによって、計算を簡易にすることができる。特にAを何回もかけるような状況の時。
- ベクトルは状態を表すことができる。また変化を線形代数で表すことができる。そうすると時系列の変化などはAの掛け算の繰り返しで表すことができる。
- 固有値分解すると、 $A^n$ 
  - を $V\Lambda^n V^{-1}$
  - とでき、 $\Lambda$
  - は対角行列なので、結局対角成分である各固有値をn乗することに等しく、計算量を大幅に削減できる。

### • 特異値分解

- 正方形以外の行列で固有値分解と似たようなことを実施する技法
- $Mv = \sigma u$
- $M^T u = \sigma v$
- となるような特殊な単位ベクトルu,vがあれば可能。
- 書き換えると
- $MV = US$

- $M^T U = V S^T$
- 変換して
- $M = U S V^T$
- $M^T = V S^T U^T$
- 掛け合わせて
- $M M^T = U S V^T V S^T U^T = U S S^T U^T$
- 特異値分解の利用例
  - 画像のデータの分解
    - 特異値を取り出すことで人間の特徴を取り出すことができる。

# 統計学

## 集合と確率

- 集合とは何か
  - ものの集まり、またそれを記述する方法
  - $S = a, b, c, d, e, f, g$   
の時
    - $a \in S$
    - $b \in S$
  - $M = c, d, g$   
の時
    - $M \subset S$
  - 集合の「要素」(または「元」) (上でいうaとかbとか) 同士は明確に区別できる。「集合」に含まれるか、含まれないかも明確に区別できる。
  - $h \notin S$   
というように集合に属さないように評価することができる。
  - 和集合と共通部分 (略)
  - 絶対補  $\bar{A}$
  - バーと相対補  $B \setminus A$   
みたいな)
- 確率
  - 確率には、一般的な頻度確率 (客観確率) とベイズ確率 (主観確率) が存在する。
    - 頻度確率は何回も測定できるものを想定 (客観的に調べることができる)
    - 確率は信念の度合いとも考えることができる。

- ベイズは客観的に確認できなくても評価できるんじゃないかという発想に面白さがある。
- もともとの全数が把握できないものについて、いろいろな条件を踏まえて、主観的に観測していくのがベイズ確率。
- 定義してしまえば、双方とも確率として数学で同様に扱うことができる。
- 確率の定義
  - $P(A) = \frac{n(A)}{n(U)} = \frac{\text{事象}A\text{が起こる数}}{\text{すべての事象の数}}$
- $P(A)$ と表すことで  $P(\bar{A}) = 1 - P(A)$   
とあらわせ、観測されていなくても、間接的に様々な確率を知ることができる。
- 共通部分  $P(A \cap B) = P(A)P(B|A)$
- 条件付き確率  $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{n(A \cap B)}{n(A)}$ 
  - 全体と条件が変わっているので注意。
- $P(A)$ と $P(B)$ が独立の場合、 $P(B|A) = P(B)$   
になるので、 $P(A \cap B) = P(A)P(B)$  と積のようになる。
- 和集合  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ 
  - $P(A \cap B)$ がかぶっている部分を引いている。
- ベイズ則
  - $P(A)P(B|A) = P(B)P(A|B)$   
を応用する。
    - $P(\text{飴玉}) = \frac{1}{4}P(\text{笑顔}|\text{飴玉}) = \frac{1}{2}P(\text{笑顔}) = \frac{1}{3}$   
の時
    - $P(\text{笑顔}, \text{飴玉}) = P(\text{飴玉}, \text{笑顔})$
    - $P(\text{飴玉}, \text{笑顔}) = P(\text{飴玉}|\text{笑顔})P(\text{笑顔})$
    - $P(\text{飴玉}|\text{笑顔}) = \frac{3}{8}$   
と計算できる。

## 統計学

- 統計
  - 記述統計と推測統計
    - データを整理しないと、我々はデータを理解できない。 →記述統計
    - 集団の性質を要約して記述する（要約統計）
    - 全数調べられるものはほとんどないので、標本を抽出（推測統計）
    - 有名なのはギネスビール？

- この講義ではビッグデータなどを扱い、そのデータを理解することが重要なため、記述統計を優先。
- 確率変数と確率分布
  - 確率変数・・・事象と結びつけられた数値（あるいは事象そのもの）例 くじの当たり額もしくは引かれたくじの番号（当たりかどうかと結びつく）
  - 確率分布・・・各事象の発生確率の分布
- 期待値
  - その分布における確率変数の平均の値もしくは「あり得そうな」値
  - 離散値の時  $E(X) = \sum P(X)f(X)$
  - 連続値の時  $E(X) = \int P(X)f(X)$
- 分散と共分散
  - 分散 データがどれだけ散らばっているか
    - $V(X) = E(X - E(X))^2 = E((X)^2) - E(X)^2$
  - 共分散 2個のデータ系列がどれだけ同じ（違う）傾向を示すか。
    - 同じ傾向ならば正の値、違う傾向なら負の値、関係性がないなら0に近い値。
    - $COV(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$
- 標準偏差
  - 分散だと単位が変わってしまうので、単位が元に戻るようルートをとったものが標準偏差
- 様々な確率分布
  - ベルヌーイ分布・・・2値の値。コイントスとか。離散分布
    - $P(X|\mu) = \mu^X * (1 - \mu)^{(1-X)}$
  - マルチヌーイ分布(カテゴリカル分布)・・・ベルヌーイの複数值取る版
  - 二項分布・・・ベルヌーイ分布の多試行版
    - $P(X|\lambda, n) = \frac{n!}{x!(n-x)!} * \lambda^x (1 - \lambda)^{n-x}$
    - サンプル十分多ければ、正規分布に近似することが知られている。
  - ガウス分布
    - 釣鐘型の連続分布
- 推定
  - 推定とは母集団を特徴付ける母数（パラメータ）を統計学的に推定すること
  - 推測統計・・・サンプルから元の母集団の姿を推定する統計領域
    - 点推定や区間推定
  - もとまった母数は推定値あるいは推定量と言われる。
    - 推定量(estimator)・・・パラメータを推定するために利用する数値の計算方法や計算式のこと。推定関数とも
    - 推定値(estimate)・・・実際の試行で計算した値

- 推定された値は $\hat{\cdot}$ が記号の上についていることが多い。
- 推定量の例
  - 標本平均
  - 特徴
    - 一致性・・・サンプル数が大きくなれば母集団に近づく。
    - 不偏性・・・サンプル数によらず、その期待値はその母集団の期待値と同様。
- 標本分散
  - 一致性はあるが、不偏性は満たさない。
  - たくさんのデータがあるときと、少数のデータのばらつきでは、少数のデータの方がばらつきが少なくなってしまう。
  - 元々の分散に近づけるために、不偏分散を利用する。
  - $\text{標本分散} \times \frac{n}{n-1}$   
 とすると不偏性を満たす。
  - 標本分散は平均が計算に入っているので、自由度が1下がってしまう。
  - そこを考慮に入れて、 $n-1$ が不偏分散の分母に来る。

## 情報理論

- 情報科学
  - 情報量が1増える場合でも、元々の量が違うと見分けやすいか見分けにくいかわかる。
  - それを数量的に判断するのが情報量
  - 具体的には増加の比率として表す。
  - 100gが110gになるときづく人は1000gが1010gになると気づくかということそうではない。比率の場合が多く、1100gとかにしないとわからないパターンが多い。
  - 自己情報量
    - 数が1ずつ増えているときに全体に対する増分の比率は 全体を $W$ (事象の総数)として  $1/W$ それを連続変数と考えて積分すると、 $\log W$ となる。これが情報量を表す単位となる。
    - $I(x) = -\log(P(x)) = \log(W(x))$
    - $P$ (確率)は $W$ (事象の総数)の逆数になるので、 $\log$ をとると-が表に出て来る。
    - 対数の底が2の時は、単位はbit
    - 自然対数の時は単位はnat
    - スイッチの数で考えると、2でログを取るという発想になる。(と考えることもできる。)
    - 確率は珍しいほど確率が低いので、逆数をとって、ログを取っているのって、結果マイナスにすると、珍しいほど情報量が大きくなるという直感に即した結果になる。

。シャノンエントロピー・・・自己情報量の期待値。

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

- シャノンエントロピーは確率 $x=0.5$ の時に最大になる傾向。逆に0 or 1になるとつまり決まった結果が出るということで、新たな情報量が少ないので、新しく得られる情報は少ないので、シャノンエントロピーの値は0になる。

。カルバックライブラーダイバージェンス

- 元々考えられていた分布 (Q) が実際測って見たら (P) だったというときにその違いはどれくらいかというときに使うのが、カルバックライブラーダイバージェンスという。
- Pという確率分布とqという確率分布の性質の違いを距離のようなものを表していると考えられる。性質は距離に近い。距離の定義をすべて満たしているわけではない。(逆から測ったら同じにならない)

- $D_{KL}(P\|Q) = E_x P(\log(\frac{P(x)}{Q(x)})) = E_x P[\log(P(x)) - \log(Q(x))] = E_x P[I(Q(x)) - I(P(x))]$

- 最初に見積もった情報量と後からわかった情報量の差の期待値(新しい分布で情報は比較される)

- カルバックライブラーダイバージェンスは情報利得と捉える考え方もできる。

- $D_{KL}(P\|Q) = E_x P(\log(\frac{P(x)}{Q(x)})) = \sum (P(x) * \log(\frac{P(x)}{Q(x)}))$

- シャノンエントロピーと形が似ている。

- 関係している。

。交差エントロピー

- KLダイバージェンスの一部分を取り出したもの
- 想定していた分布Qについての自己情報量 $I(Q)$ を分布Pで平均をとる。

- $D_{KL}(P\|Q) = \sum (P(x) * ((-\log(Q(x))) - (-\log(P(x))))$   
における

- $\sum (P(x) * -\log(Q(x)))$   
の部分が交差エントロピー $-H(P,Q)$ で表す

- $\sum (P(x) * -\log(P(x)))$   
を $H(P)$ として、

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

- また

- $H(P, Q) = -E_x P[I(Q(x))]$