応用数学・要点まとめ

第１章 線形代数

▣ 逆行列

元の行列 A との積が単位行列 I になる行列 A^{-1}：　$AA^{-1}=A^{-1}A=I$

逆行列 A^{-1} は、掃き出し法（左右の係数の行列に同じ基本変形を実行）によって求める：

$$\left(\begin{array}{cc|cc} 1 & 4 & 1 & 0 \\ 2 & 6 & 0 & 1 \end{array}\right) \xrightarrow[\text{2 行目を 1/2 倍}]{} \left(\begin{array}{cc|cc} 1 & 4 & 1 & 0 \\ 1 & 3 & 0 & 1/2 \end{array}\right) \xrightarrow[\text{1 行目に 2 行目の -1 倍を加える}]{} \left(\begin{array}{cc|cc} 0 & 1 & 1 & -1/2 \\ 1 & 3 & 0 & 1/2 \end{array}\right)$$

$$\xrightarrow[\text{2 行目に 1 行目の -3 倍を加える}]{} \left(\begin{array}{cc|cc} 0 & 1 & 1 & -1/2 \\ 1 & 0 & -3 & 2 \end{array}\right) \xrightarrow[\text{1 行目と 2 行目を入れ替える}]{} \left(\begin{array}{cc|c} 1 & 0 & \boxed{\begin{matrix} -3 & 2 \\ 0 & -1/2 \end{matrix}} \\ 0 & 1 & \end{array}\right)$$

逆行列

▣ 行列式

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad-bc$$

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

行列式の特徴：

①同じ行ベクトルが含まれると行列式は 0

②1 つのベクトルが λ 倍されると行列式は λ 倍される

③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、行列式の足し合わせになる

④行を入れ替えると符号が変わる

← 入れ替えた前後の行列を加えると操作した 2 つの行が同じになり、行列式が 0 になるため

▣ 固有値分解

$A\vec{x}=\lambda\vec{x}$　　λ：固有値、$\vec{x}$：固有ベクトル ⇒ **主成分分析**による**次元削減**で利用

$|A-\lambda I| = 0$　固有値が存在する条件

$A=\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}$ の場合、$\begin{vmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{vmatrix}=0$ より、$(1-\lambda)(3-\lambda)-8=0$ → $\lambda=5$ or -1

λ=5 の時、$\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} x1 \\ x2 \end{pmatrix}=5\begin{pmatrix} x1 \\ x2 \end{pmatrix}$ より、$x_1=x_2$ → $\vec{x}=\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ の定数倍

λ=-1 の時、$\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} x1 \\ x2 \end{pmatrix}=-1\begin{pmatrix} x1 \\ x2 \end{pmatrix}$ より、$x_1=-2x_2$ → $\vec{x}=\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ の定数倍

⇔ $AV = V\Lambda$ → $A = V\Lambda V^{-1}$　　ここで、A：正方行列、$\Lambda=\begin{pmatrix} \lambda 1 & & \\ & \lambda 2 & \\ & & \ddots \end{pmatrix}$、$V=\begin{pmatrix} \vec{v1} & \vec{v2} & \cdots \end{pmatrix}$

▣ 特異値分解

正方行列以外は固有値分解できないため、その拡張として特異値分解を用いる ⇒ **次元削減**で利用

$M\vec{v} = \sigma\vec{u}$ 、$M\vec{u} = \sigma\vec{v}$ となる特殊な単位ベクトルがある ⇒ $M = USV^{-1}$ **特異値分解**

$MV=US \rightarrow M=USV^{-1}$ 、$M^TU=VS^T \rightarrow M^T = VS^TU^{-1}$

これらの積　$MM^T=USS^TU^{-1}$、$M^TM=VS^TSV^{-1}$

応用数学・要点まとめ

MM^T (M^TM) を固有値分解すれば、その左特異ベクトル U、右特異ベクトル V (単位ベクトル)
と特異値の2乗が求められる

第２章 確率・統計

▣ 確率

- 確率の種類：
  頻度確率 (客観確率)：発生する頻度
  ベイズ確率 (主観確率)：信念の度合い
- 条件付き確率： ある事象 X=x が与えられた下で、Y=y となる確率
  P (Y = y | X = x) ＝P (Y = y, X = x) ／P (X = x)
- 独立な事象の同時確率
  P (X = x, Y = y) ＝P (X = x) P (Y = y) ＝P (Y = y, X = x)
- **ベイズ則**： 一般的に事象 X＝x と事象 Y＝y に対して以下が成立
  P (X = x | Y = y) P (Y = y) ＝P (Y = y | X = x) P (X = x)

▣ 確率変数と確率分布

- **確率変数** ：事象と結び付けられた数値 (事象そのものを指すと解釈する場合も多い)
- **確率分布** ：事象の発生する確率の分布 (拡散値であれば表に示せる)

▣ 期待値と分散・共分散

- **期待値** ： $E(f) = \sum_{k=1}^{n} P(X = x_k)\, f(X = x_k)$ ・・・離散値の場合
  $E(f) = \int P(X = x)\, f(X = x)\, dx$ ・・・連続値の場合
- **分散** データの散らばり具合 (期待値からのズレの平均)：
  $Var(f) = E[(f(X=x) - E(f))^2] = E(f^2(X=x)) - (E(f))^2$
- **共分散** 2つのデータ系列の傾向の違い (正は似た傾向、負は逆の方向、0 は関係性に乏しい)：
  $Cov(f, g) = E[(f(X=x) - E(f))(g(Y=y) - E(g))] = E(fg) - E(f)E(g)$

▣ 様々な確率分布

- **ベルヌーイ分布** ： コイントスのイメージ (裏と表で出る割合が等しくなくても扱える)
  $P(X \mid \mu) = \mu^x (1-\mu)^{1-x}$
- **マルチヌーイ (カテゴリカル) 分布** ： さいころを転がすイメージ (各面の出る割合が等しくなくても扱える)
- **二項分布** ： ベルヌーイ分布の多試行版
  $P(X \mid \lambda, n) = \binom{n}{x} \lambda^x (1-\lambda)^{n-x}$
- **ガウス分布** ： 釣鐘型の連続分布
  $$N(x; \mu, \sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

第３章 情報理論

▣ 自己情報量

応用数学・要点まとめ

情報量： ある事象が起きた際、それがどれほど起こりにくいかを表す尺度。
その事象が本質的にどの程度の情報を持つかの尺度とみなせる（有用性とは無関係）。

自己情報量： それぞれの事象の情報量

ある事象 X が起こる確率 P(x)、情報の変化率を W(x) とすると、

$I(x)=-\log_s(P(x))=\log_s(W(x))$

S=2 の時はビット(bit)、S=e（ネイピア）の時はナット(nat)

シャノンエントロピー： 自己情報量の期待値

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

▣ カルバック・ライブラー・ダイバージェンス（Kullback-Leibler divergence、KL 情報量）

- **カルバック・ライブラー・ダイバージェンス** ： 2つの確率分布がどの程度似ているかを表す尺度

$$D_{KL}(P\|Q)=E_{X\sim P}\left[\log\frac{P(x)}{Q(x)}\right]=E_{X\sim P}[\log P(x)-\log Q(x)]=\sum_x P(x)\log\frac{P(x)}{Q(x)}$$

- **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、Q についての自己情報量を P の分布で平均

$$H(P,Q)=H(P)+D_{KL}(P\|Q)=-E_{X\sim P}[\log Q(x)]=-\sum P(x)\log(Q(x))$$