実装演習レポート

科目：応用数学

【第一章：線形代数】

要点

[行列]

スカラーは普通の数であり、演算が可能である。

ベクトルは大きさと向きを持ち、スカラーのセットで表示される。

行列は、スカラーを表にしたものやベクトルを並べたものであるといえる。行列の積は連立方程式の研究から生まれた概念である。

逆行列は、逆数のようなものであり、ある行列とその逆行列をかけると単位行列となる。逆行列は行基本変形を記録して行列と対応させて掛け算していくと求まり、これを掃き出し法と呼ぶ。

$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$という行列において、ad-bc=0 の時、逆行列は存在しない。

$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \overrightarrow{v_1} \\ \overrightarrow{v_2} \end{pmatrix}$ で作られる平行四辺形の面積を$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} \overrightarrow{v_1} \\ \overrightarrow{v_2} \end{vmatrix}$と表し、行列式と呼ぶ。

n 個のベクトルから構成される行列式は以下のような特徴を持つ。

同じ行ベクトルが含まれていると行列式は 0 になる。

1 つのベクトルがλ倍されると行列式はλ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は、行列式の足し合わせになる。

行列式は下記のように求められる。

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

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

[固有値と固有ベクトル]

ある行列 A に対して、以下のような式が成り立つ特殊なベクトル$\vec{x}$と係数λを、行列 A に対する固有ベクトル、固有値という。

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

固有値・固有ベクトルの求め方は以下の通りである。

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

$$(A - \lambda I)\vec{x} = \vec{0}$$

$$\vec{x} \neq \vec{0}$$

より

$$|A - \lambda I| = 0$$

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} \text{の時}$$

$$\begin{vmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{vmatrix} = 0$$

$$(1-\lambda)(3-\lambda) - 4 \cdot 2 = 0$$

$$\lambda = 5 \text{ or } -1$$

固有値分解とは、正方形の行列を以下のような 3 つの行列の積に変換することである。

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

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \end{pmatrix}、V = (\overrightarrow{v_1}\ \overrightarrow{v_2}\ \cdots)$$

固有値 $\lambda_1, \lambda_2, \cdots,$　固有ベクトル $\overrightarrow{v_1}, \overrightarrow{v_2}, \cdots,$

特異値分解とは、正方行列以外で特殊な単位ベクトルが存在する際の固有値分解のようなものである。

$M\vec{v} = \sigma\vec{u}$　　　　σ：特異値、$\vec{u}$：特異ベクトル（単位ベクトル）

$M^T\vec{u} = \sigma\vec{v}$

$$M = USV^{-1}$$

特異値の求め方

$MV = US$　　　　$M^TU = VS^T$

$M = USV^{-1}$　　　　$M^T = VS^TU^{-1}$

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

MM^T を固有値分解すれば、その左特異ベクトルと特異値の 2 乗が求められる。

【第二章：確率・統計】

<u>要点</u>

[確率]

事象が発生する頻度を頻度確率（客観確率）といい、信念の度合いをベイズ確率（主観確率）という。

[条件付き確率]

ある事象 X=x が与えられた下で、Y=y となる確率を条件付き確率という。

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

[独立な事象の同時確率]

お互いの発生には因果関係のない事象 X=x と事象 Y=y が同時に発生する確率は以下のように表す。

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

[ベイズ則]

一般に事象 X=x と事象 Y=y に対して、下記の式が成り立つ。

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

1/4 の確率で生じるある事象 X が発生しているときの Y が発生する確率が 1/2 の時、Y が発生しているときの X が発生している確率を求める。（ただし、Y の発生確率は 1/3 である。）

P(X) = 1/4, P(Y|X) = 1/2, P(Y)=1/3

P(Y|X) x P(X) = P(Y, X) : 1/2 x 1/4 = 1/8

P(Y, X) = P(X, Y)

P(X, Y) = P(X|Y) x P(Y) : 1/8 = P(X|Y) x 1/3

よって、P(X|Y) = 3/8

[確率変数と確率分布]

確率変数は事象と結び付けられた数値であり、事象そのものを指す場合が多い。確率分布は事象の発生する確率の分布であり、離散値であれば表に示すことができる。

事象	裏が 0 枚、表が 4 枚	裏が 1 枚、表が 3 枚	裏が 2 枚、表が 2 枚	裏が 3 枚、表が 1 枚	裏が 4 枚、表が 0 枚

確率変数（裏を 0，表を 1 と対応させ和をとった）	4	3	2	1	0
事象が発生した回数	75	300	450	300	75
事象と対応する確率	1/16	4/16	6/16	4/16	1/16

[期待値]

その分布における、確率変数の　平均の値 or 「ありえそう」な値

事象 X	X1	X2	・・・	Xn
確率変数 f(X)	f(X1)	F(X2)	・・・	f(Xn)
確率 P(X)	P(X1)	P(X2)	・・・	P(Xn)

期待値 E(f)

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

連続する値なら・・・

期待値 E(f)

$$= \int P(X = x) f(X = x) dx$$

[分散と共分散]

分散はデータの散らばり具合を示し、データの各々の値が、期待値からどれだけズレているのか平均したものである。

分散 Var(f)

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

共分散は 2 つのデータ系列の傾向の違いを示し、正の値の場合は似た傾向を意味し、負の値の場合は逆の傾向を意味する。ゼロの場合は関係性に乏しいということになる。

共分散 Cov(f, g)

$$= E((f_{(X=x)} - E(f))(g_{(Y=y)} - E(g))$$
$$= E(fg) - E(f)E(g)$$

[分散と標準偏差]
分散は 2 乗してしまっているので元のデータと単位が違うことになる。
2 乗することの逆演算（つまり平方根を求める）をすれば元の単位に戻る。

$$\begin{aligned}\sigma &= \sqrt{Var(f)} \\ &= \sqrt{E((f_{(X=x)} - E_{(f)})^2})\end{aligned}$$

[様々な確率分布 I]
ベルヌーイ分布はコイントスのイメージであり、裏と表で出る割合が等しくなくとも扱える。

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

マルチヌーイ（カテゴリカル）分布はサイコロを転がすイメージであり、各面の出る割合が等しくなくとも扱える。

[様々な確率分布 II]
二項分布はベルヌーイ分布の多試行版である。

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

ガウス分布は釣鐘型の連続分布である。

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

【第三章：情報理論】
[自己情報量]
対数の底が 2 のとき、単位はビット(bit)であり、対数の底がネイピアの e のとき、単位は(nat)である。

$$I(x) = -\log\bigl(P(x)\bigr) = \log\,(W(x))$$

[シャノンエントロピー]
微分エントロピーともいうが、微分しているわけではない。
自己情報量の期待値

$$H(x) = E\bigl(I(x)\bigr) = -E\Bigl(\log\bigl(P(x)\bigr)\Bigr) = -\sum\Bigl(P(x)\log\bigl(P(x)\bigr)\Bigr)$$

[カルバック・ライブラー　ダイバージェンス]
同じ事象・確率変数における異なる確率分布 P,Q の違いを表す。

$$E\left(f(x)\right) = \sum_x P(x)f(x)$$

$$D_{KL}(P \parallel Q) = \mathbb{E}_{x\sim P}\left[log\frac{P(x)}{Q(x)}\right] = \mathbb{E}_{x\sim P}[logP(x) - logQ(x)]$$

$$I\left(Q(x)\right) - I\left(P(x)\right) = \left(-\log\left(Q(x)\right)\right) - \left(-\log\left(P(x)\right)\right) = log\frac{P(x)}{Q(x)}$$

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

[交差エントロピー]
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 の分布で平均している。

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

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

$$H(P,Q) = -\mathbb{E}_{x\sim P}logQ(x) = -\sum_x P(x)logQ(x)$$