第一章　線形代数

１）固有値・固有ベクトル

- ある行列 $\boldsymbol{A}$ に対して、式$\boldsymbol{Ax} = \lambda\boldsymbol{x}$が成り立つとき、$\lambda$を固有値、$\mathbf{0}$でないベクトル$\boldsymbol{x}$を固有ベクトルという。

問題

$\begin{bmatrix} 3 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$の固有値、固有ベクトルを求める

回答

$\boldsymbol{(A - \lambda I)x = 0}$、$\boldsymbol{x \neq 0}$より

$|\boldsymbol{(A - \lambda I)}|$ = 0

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

$(3-\lambda)(2-\lambda)(1-\lambda) = 0$　よって、$\lambda = 1$ or 2 or 3

$\lambda = 1$のとき、下記を計算して$x_1 = x_2 = 0$

- $2 \cdot\ x_1 + 2 \cdot\ x_2 = 0$
- $1 \cdot\ x_2 = 0$
- $0 \cdot\ x_3 = 0$

よって、$\lambda = 1$のときの固有ベクトルは、$t \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ (tは0以外の定数)

$\lambda = 2$のとき、下記を計算して$x_1 = -2x_2, x_3 = 0$

- $1 \cdot\ x_1 + 2 \cdot\ x_2 = 0$
- $0 \cdot\ x_2 = 0$
- $-1 \cdot\ x_3 = 0$

よって、$\lambda = 2$のときの固有ベクトルは、$t \cdot \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$ (tは0以外の定数)

$\lambda = 3$のとき、下記を計算して$x_2 = x_3 = 0$

- $0\cdot\ x_1 + 2\cdot\ x_2 = 0$
- $-1\cdot\ x_2 = 0$
- $-2\cdot\ x_3 = 0$

よって、$\lambda = 3$のときの固有ベクトルは、$t\cdot\ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ (tは0以外の定数)

２）固有値分解

$\boldsymbol{\lambda, x}$ を行列$\boldsymbol{A}$の固有値、固有ベクトルとする。
λを大きさの順で並べ替え、これを対角にする対角ベクトルΛをつくる。
さらに、対応する$\boldsymbol{x}$を同順で列にして行列$\boldsymbol{V}$をつくる。
すると、

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

となる行列$\boldsymbol{V}$を取ることができる。
これを行列$\boldsymbol{A}$の固有値分解という。

３）特異値・特異値分解

$\boldsymbol{M}$の特異値分解は下記のように表される。

$$\boldsymbol{M} = \boldsymbol{USU^T}$$

第二章　確率・統計

１）条件付き確率

ある事象Bという条件のもとで事象Aとなる確率

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

２）ベイズ則

事象$X = x$と事象$Y = y$に対して、

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

３）期待値・分散

- 期待値$E(f)$
  $= \sum_{k=1}^{n} P(X - x_k)f(X = x_k)$
- 分散$Var(f)$

$$= E((f_{X=x} - E_f)^2)$$
$$= E(f^2_{(X=x)}) - (E(_f))^2$$

第三章　情報理論

1）自己情報量・シャノンエントロピー

- シャノンエントロピーは、事象Eが起こる確率を$P(E)$とし、すべての事象$E \in \Omega$に対して情報量の期待値のこと
  $H = -\sum_{E \in \Omega} P(E) log P(E)$

2）KLダイバージェンス・交差エントロピー

- 交差エントロピーは2つの確率分布の間に定義される尺度
  $H(p, q) = -q(x) ln P(x)$
- KLダイバージェンスは2つの確率分布の「差異」を計る尺度
  $KL[q(x)||p(x)] = Q(x) ln \frac{P(x)}{Q(x)} = Q(x) ln P(x) - Q(x) ln Q(x)$