応用数学

第一章　線形代数

かけてもかけられても変化しない下記Iのような行列を単位行列という。

$$I = \begin{pmatrix} 1 & & \\ & 1 & \\ & & \ldots \end{pmatrix}$$

ある行列Aに対して$AA^{-1} = I$となるようなA^{-1}を、Aの逆行列という。

$$\left(\begin{matrix} a & b \\ c & d \end{matrix}\middle|\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}\right) ・・・①$$

式①のようにある行列と単位行列を並べ、左側の行列が単位行列になるように変形すると、右側に逆行列が完成する。このような逆行列の求め方を掃き出し法という。
解が一つに定まらないような連立方程式の係数を抜き出したような行列には逆行列は存在しない。

ある行列が二つのベクトルだと仮定した時に作られる平行四辺形の面積のことをその行列の行列式という。

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc ・・・②$$

2×2 の正方行列の行列式は式②のように求めることができる。

$$A\vec{x} = \lambda\vec{x} ・・・③$$

ある行列Aが存在するとき、式③が成り立つようなベクトル$\vec{x}$と係数λが存在する。このベクトル$\vec{x}$を行列Aの固有ベクトル、係数λを行列Aの固有値という。

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

のとき、Aの固有値は

$$\begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = 0$$

のλにおける方程式の解である。
固有値をλ_1, λ_2としたときAの固有ベクトルは

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda_2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

のx_1, x_2における連立方程式の解である。

固有値$\lambda_1, \lambda_2 \dots$と固有ベクトル$x_1, x_2 \dots$を持つ正方行列を$A$、固有値を対角線上に並べそれ以外の値を 0 とした行列をΛ、対応する固有ベクトルを並べた行列をVとしたとき、

$$AV = V\Lambda$$

が成り立ち、変形すると

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

が成り立つ。このような正方行列を 3 つの行列の積に変換することを固有値分解という。

正方行列でない行列は固有値分解できないが、転置行列との積を固有値分解することで特異ベクトルと特異値を求めることができる。これを特異値分解という。特異値分解は画像のデータ量を小さくする際に用いることができる。

第二章 確率・統計

ある事象与えられた状態での別の事象の発生確率を条件付き確率という。条件付き確率は事前に与えられた事象がX、後に起こる事象をYとすると、

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

で求めることができる。
互いに独立である 2 つの事象が同時に発生する確率は、それぞれの確率の積で求めることができる。

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad ・・・①$$

式①をベイズの定理という。
投げたコインの表の枚数や、サイコロの目などの事象に結びつけられた値を確率変数という。事象が発生する確率の分布のことを確率分布という。
ある事象の確率変数と確率との積を期待値という。
あるデータ群がどれだけ散らばっているかを表す値に分散がある。分散はそれぞれのデータと期待値との差を二乗したものの平均である。
二つのデータ群の傾向を表す値に共分散がある。共分散はそれぞれのデータの積の期待値からそれぞれの期待値の積を引いたものであり、正の値であれば似た傾向、負の値であれば逆の傾向があるといえる。
分散は求める過程で二乗してしまっているため、元のデータと単位が異なってしまう。この問題を解決するために分散の平方根を取ったものを標準偏差という。
試行結果が 2 種類しかないような確率分布をベルヌーイ分布という。ベルヌーイ分布は 1 である確率がpのとき、0 である確率が$p-1$となる。
ベルヌーイ分布の試行結果が 2 種類であるのに対し、多種類ある場合の確率分布をマルチヌーイ分布という。
ベルヌーイ分布は試行回数が 1 回であるのに対し、試行回数を増やした確率分布を二項分布という。
他にも代表的な確率分布に、縦軸に確率、横軸に確率変数を取った際に釣鐘型の連続分布が出来上がるガウス分布がある。

第三章 情報理論

ある事象が状態の不確定さをどれだけ減少させるかを自己情報量という。確率が小さい事象ほど自己情報量は大きくなるといえる。

$$I(x) = -\log(\mathrm{P}(x)) = \log(W(x))$$

自己情報量の期待値をシャノンエントロピーという。

$$H(x) = E\big(I(x)\big)$$

ある確率分布についての情報量を別の確率分布で平均したものを交差エントロピーという。

二つの確率分布の差異を数量化したものをカルバック・ライブラーダイバージェンスという。交差エントロピーから情報エントロピーを引くことで求められる。二つの確率分布が近ければ値は 0 に近づいていく。