応用数学

「第1章：線形代数」

・ベクトルの変換　＞　行列とベクトルの積で行う

・行列とベクトルの積　　行列（横）× ベクトル（縦）

計算結果はベクトルの数と同じになる。

・固有値、固有ベクトル

複雑なベクトル計算がある特定の条件下では　$A\vec{x} = \lambda\vec{x}$

で求められる。

この特殊なベクトルXとその係数λを、行列Aに対する固有ベクトル、固有値という。

・行列式～平行四辺形の面積と考えられる

複雑な行列～立面体の面積

・固有値分解～計算が楽になる。大きいデータをちいさくできる。

データを整理して表現できる。

行列Aがλ1,λ2,λ3をとるとき、このλnを対角線に並べた行列を作る。

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \end{pmatrix}$$

これに固有ベクトル　$V = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \cdots \end{pmatrix}$

を用意すると　$AV = V\Lambda$

となる。左辺からVを消すためVの逆行列を掛ける。

このように正方形の行列を３つの行列の積に分解することを固有値分解という。

・特異値分解～正方行列以外（長方形）は固有値分解ではなく特異値分解で対応。

かけ算して正方行列を作り固有値分解の手法で解く。

特長の見えない行列の特長を見つけるのが特異値分解。

音声データや画像データで同じ特長を抽出したいときに特異値分解を使用。

「第2章：確率・統計」

・頻度確率～通常の確率

・ベイズ確率～信念の度合の確率。主観確率。

・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

$$P(Y=y \mid X=x) = P(Y=y, X=x) / P(X=x)$$

・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。

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

・ベイズ則～事象Aが起こる条件のもとで、事象Bが起こる確率。

$$P(X=x \mid Y=y)P(Y=y) = P(Y=y \mid X=x)P(X=x)$$

・確率変数～事象と結びつけられた数値

・確率分布～事象の発生する確率の分布

・期待値～確率変数の平均値、またはありえそうな値

離散値：

$$\text{期待値}\mathrm{E}(f) = \sum_{k=1}^{n} P(X = x_k)f(X = x_k)$$

連続値：

$$\text{期待値}\mathrm{E}(f) = \int P(X = x)f(X = x)dx$$

・分散～データの各々の値が平均値からどれだけズレているかを表したもの

$$\begin{aligned}&\text{分散}\mathrm{Var}(f)\\ &= \mathrm{E}\left(\left(f_{(X=x)} - \mathrm{E}_{(f)}\right)^2\right)\\ &= \mathrm{E}\left(f^2_{(X=x)}\right) - \left(\mathrm{E}_{(f)}\right)^2\end{aligned}$$

・共分散～２つのデータの系列の違い。正の値を取れば似た傾向。負の値を取れば似た傾向。ゼロをとれば関係性は乏しい。

$$\begin{aligned}&\text{共分散}\mathrm{Cov}(f, g)\\ &= \mathrm{E}\left(\left(f_{(X=x)} - \mathrm{E}(f)\right)\left(g_{(Y=y)} - \mathrm{E}(g)\right)\right)\\ &= \mathrm{E}(fg) - \mathrm{E}(f)\mathrm{E}(g)\end{aligned}$$

・ベルヌーイ分布～コイントスのイメージ。２つしか値をとらない場合の分布。離散値の確率分布。

$$P(x \mid \lambda) = \lambda^x (1-\lambda)^{(1-x)} \text{ for } x \in \{0, 1\}$$

Takes a single parameter $\lambda \in [0, 1]$

・マルチヌーイ分布〜ベルヌーイ分布を一般化した離散値の確率分布。

２つではなくN個の値をとる。

$$P(\boldsymbol{x} \mid \boldsymbol{\lambda}) = \prod_{i=0}^{K} \lambda_i^{x_i}$$

Takes K parameters $\lambda_i \in [0,1]$ where $\sum_{i=1}^{K} \lambda_i = 1$

・二項分布〜ベルヌーイ分布の多試行版。

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

・ガウス分布〜釣鐘型の連続分布。

最大値をとるまで緩やかに上昇し、最大値を超えると緩やかに下降。

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

exp(x) = ネイピア関数、微分しても値が変わらない。

「第3章：情報理論」

・自己情報量〜対数の底が２のとき単位はbit

公式のWは逆数、事象の数。e.g.) log(1/A) = -logA

logは桁数の多い掛算を足し算で演算できるメリットがある。

e.g) log2(A×B) = log2A + log2B

$$I(x) = -\log\big(P(x)\big) = \log\big(W(x)\big)$$

対数の底がネイピアのとき単位はnat

・シャノンエントロピー〜平均情報量。自己情報量の期待値。自己エントピーとは別物。

$$H(P) = -\sum_{A \in \Omega} P(A) \log P(A)$$

・KL ダイバージェンス〜同じ事象・確率変数の異なる確率分布P、Qの違いを表す。

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

・交差エントロピー〜KLダイバージェンスの一部を取り出し、PとQの平均を考える。

Qの自己情報量をPの分布で平均する。

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

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