応用数学レポート

第1章 線形代数

本章で学ぶ項目は以下の通り。

1. 固有値・固有ベクトルの求め方
2. 固有値分解とは
3. 特異値・特異ベクトルの概要
4. 特異値分解の概要

準備として、スカラー、ベクトル、行列、行列積についても確認。

1.1 固有値・固有ベクトル

ある正方行列Aに対して $A\mathbf{x} = \lambda\mathbf{x}$ を満たすベクトル $\mathbf{x}$ を固有ベクトル、λ を固有値という。

固有方程式 $|A - \lambda I| = 0$ を解くことで固有値を求め、元の方程式に固有値を代入することで固有ベクトルを得る。

1.2 固有値分解

固有値を対角線状に並べた対角行列を Λ 、対応する固有ベクトルを並べた行列を V としたとき、$AV = V\Lambda$ と関係付けられる。$VV^{-1} = I$ であるとき、$A = V\Lambda V^{-1}$ と変形できる。この変形を固有値分解と呼ぶ。

1.3 特異値・特異ベクトル・特異値分解

行列 M に対して

$M\mathbf{v} = \sigma\mathbf{u}$ $M^{\mathrm{T}}\mathbf{u} = \sigma\mathbf{v}$

が成り立つとき、$\mathbf{u}$ を左特異ベクトル、$\mathbf{v}$ を右特異ベクトル、σ を特異値という。左特異ベクトルを並べた行列を U、右特異ベクトルを並べた行列を V、特異値を対角成分に並べた行列を S とすると、$M = USV^{-1}$ のように変形できる。この変形を特異値分解という。

なお、$M^{\mathrm{T}}U = VS^{\mathrm{T}}$ から、$M^{\mathrm{T}} = VS^{\mathrm{T}}U^{-1}$ を得られることから、$MM^{\mathrm{T}} = USV^{-1}VS^{\mathrm{T}}U^{-1} = USS^{\mathrm{T}}U^{-1}$ となる。つまり、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二乗が求められる。同じく、$M^{\mathrm{T}}M$ を固有値分解することで右特異ベクトルが得られる。

第2章 確率・統計

本章で学ぶ内容は以下の通り。

1. 条件付き確率
2. ベイズ則の概要
3. 期待値・分散の求め方
4. 様々な確率分布の概要

2.1 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。

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

ただし、$P(y, x)$ は事象 x、y に対する同時確率。

2.2 ベイズ即

条件付き確率の定義より、

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

2.3 期待値・分散

$k = 1, ..., n$ のとき、事象 x_k、確率変数 $f(x_k)$、確率 $P(x_k)$ とすると、f を離散値としたときの期待値 $E[f]$ は、

$$E[f] = \sum_{k=1}^{n} P(x_k)f(x_k)$$

となる。連続値の場合は、

$$E[f] = \int P(x)f(x)$$

となる。分散 $\mathrm{Var}[f]$ および共分散 $\mathrm{Cov}[f, g]$ は

$$\mathrm{Var}[f] = E[f(x) - E[f(x)]^2] = E[f(x)^2] - E[f(x)]^2$$
$$\mathrm{Cov}[f] = E[(f(x) - E[f(x)])(g(x) - E[g(x)])] = E[fg] - E[f]E[g]$$

2.4 確率分布

2.4.1 ベルヌーイ分布

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

2.4.2 二項分布

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

2.4.3 ガウス分布

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

第3章 情報理論

本性で学ぶ項目は以下の通り。

1. 自己情報量・シャノンエントロピー
2. KLダイバージェンス・交差エントロピー

3.1 自己情報量

確率分布 $P(x)$ の負の対数を自己情報量という。

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

対数の底が2のとき、単位はbit、ネイピア数のとき、単位はnatという。

3.2 シャノンエントロピー

自己情報量の期待値をシャノンエントロピー（微分エントロピー）という。

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

3.3 KL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 の違いを表す量としてカルバック・ライブラー（KL）ダイバージェンスがある。

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

つまり、自己情報量の差の期待値である。

3.4 交差エントロピー

KLダイバージェンスに P のシャノンエントロピーを足したものを交差エントロピーという。

$$H(P, Q) = H(P) + D_{\mathrm{KL}}(P||Q) = \mathbb{E}_{x \sim P}[\log Q(x)]$$

Q についての自己情報量を P の分布で平均している。