線形代数

行列式

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$
$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

固有値、固有ベクトル

- ある行列Aに対してベクトルxと係数λがある
- このλを固有ベクトル、固有値という

$$A\vec{x} = \lambda\vec{x}$$
$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = 0$$

固有値分解

- 行列の累乗の計算を容易にするため
- ある正方形の行列を、固有ベクトルを並べた行列と固有値を対角線上に並べた行列（対角線以外は０にして固有値を表現）で構成される式に表現しなおす

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots \\ 0 & \lambda_2 & \dots \\ \dots & \dots & \dots \end{pmatrix}$$
$$V = \begin{pmatrix} \vec{v_1} & \vec{v_2} & \dots \end{pmatrix}$$
$$AV = V\Lambda \text{より}$$
$$A = V\Lambda V^{-1}$$

特異値分解

- 正方形以外の行列を疑似的に固有値分解すること
- 対象の行列を転置させ、転置前の行列と転置後の行列の積で正方形化する。それを固有値分解

$$MV = US \qquad M^TU = VS^T$$
$$M = USV^{-1} \qquad M^T = VS^TU^{-1}$$
$$MM^T = USV^{-1}VS^TU^{-1} = USS^TU^{-1}$$

- よって左特異ベクトルと特異値の二乗が求められる

- データを減らすとき、単純にデータを減らしていくと、元の画像と異なる画像になる
- 特異値を減らすことで、元の画像が分かるまま、データを減らすことが可能

確率・統計

頻度確率(客観確率)

- 発生する頻度

ベイズ確率(主観確率)

- 信念の度合い

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

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

独立な事象の同時確率

- お互い因果関係のない事象X,Yが同時に発生する確率

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

ベイズ則

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

期待値

- Pは確率、fは確率変数

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

- 連続する値の場合

$$\text{期待値}E(f)=\int P(x)f(x)dx$$

分散

- データの散らばり具合
- データの格々の値が期待値からどれだけずれているか平均したもの

$$分散Var(f) = E\left((f(x) - E(f))^2\right) = E(f(x)^2) - E(f)^2$$

共分散

- 2つのデータ系統の傾向のちがい
- 正(負)の値を取れば似た(逆の)傾向
- ゼロを取れば関係性とに乏しい

$$\begin{aligned}共分散Var(f) &= E\left((f(x) - E(f))(g(Y) - E(g))\right) \\ &= E(fg) - E(f)E(g)\end{aligned}$$

ベルヌーイ分布

- コイントスのイメージ
- 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

- サイコロを転がすイメージ
- 各面の出る割合が等しくなくとも扱える

二項分布

- ベルヌーイ分布の多試行版

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

ガウス分布

- 釣鐘型の連続分布

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

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)
- 対数の底がネイピアのeのとき、単位は(nat)
- 情報量→珍しいほど価値がある
- ある事象が起きた時に、どのくらいその事象が起こりにくいか

- 確率が低いほど、情報量が多い

$$I(x) = -log(P(x)) = log(W(x))$$
$$W(x) = \frac{1}{P(x)}$$

シェノンエントロピー

- 自己情報量の期待値

$$H(x) = E(I(x))$$
$$= -E(log(P(x))$$
$$= -\Sigma(P(x)log(P(x)))$$

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

$$D_{kl}(P||Q) = E_{x \sim P}\left[log\frac{P(x)}{Q(x)}\right]$$
$$= E_{x \sim P}\left[logP(x) - logQ(x)\right]$$

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

$$H(P,Q) = H(P) + D_{kl}(P||Q)$$
$$H(P,Q) = -E_{x \sim P}\ logQ(x)$$

$$-\sum P(x)\log\Big(P(x)\Big)$$