ID: vcx23527
氏名: 稲垣英樹

科目: 応用数学
要点まとめ

第1章:線形代数
機械学習のモデルを扱う上で、行列を用いた演算を多用することになるが、その手法として「逆行列」、「固有値・固有ベクトル」がある。

「逆行列」は行列の逆数のようなもので、与えられたn次正方行列 A に対して右から掛けても左から掛けても単位行列Eとなるような行列をAの逆行列といい、A^{-1} で表わす。なお、逆行列は,正方行列に対してのみ定義でき、正方行列でない行列に対する逆行列は考えない。また行列式が 0 とならない行列に対してのみ逆行列が存在する。逆行列を求める手法として「掃き出し法」がある。

また、n次正方行列Aに対して $A\vec{x}=\lambda\vec{x}$ $(\vec{x}\neq\vec{0})$ となるような定数λとベクトル$\vec{x}$(n次元の列ベクトル)が存在するとき、λをAの「固有値」といい、$\vec{x}$をλに対する「固有ベクトル」という。

正方行列 A が固有値λ、固有ベクトル$\vec{v}$を持つとき、$A=V\Lambda V^{-1}$と変形することを「固有値分解」という。

ただし、$\Lambda=\begin{pmatrix}\lambda1 & & \\ & \lambda2 & \\ & & \dots\end{pmatrix}$ $V=(\overrightarrow{v1}\ \overrightarrow{v2}\ \cdots)$

この変換によって行列の累乗の計算が容易になる等の利点がある。

正方行列以外で固有値分解を行うことを「特異値分解」といい、

$$M\vec{v} = \sigma\vec{u}$$
$$MT\vec{u} = \sigma\vec{v}$$

のような単位ベクトルが存在する場合は、以下のように特異値分解できる。

$$M = USV^{-1}$$

「固有値分解」、「特異値分解」により画像データを小さくすることで、計算量の圧縮の効果が期待できる。

第2章:確率・統計

「確率・統計」は、分類や推定、予測など、機械学習のさまざまな手法の基礎をなすものであり、その中でも以下、「条件付き確率」、「ベイズ則」、「確率分布」について、まとめる。

条件付き確率とは、「雨が降っている条件下で交通事故に逢う確率」というように、ある事象 X=x が与えられた下で、Y=y となる確率のことで、以下のようにあらわされる。このとき、事象 X=x と事象 Y=y はお互いの発生に因果関係はないものとする。

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

ベイズ確率は、事象 X=x と事象 Y=y に対し、次のようにあらわされる。

$$\mathrm{P}(\mathrm{X}=\mathrm{x} \mid \mathrm{Y}=\mathrm{y}) = \frac{\mathrm{P}(\mathrm{Y}=\mathrm{y} \mid \mathrm{X}=\mathrm{x})\mathrm{P}(\mathrm{X}=\mathrm{x})}{\mathrm{P}(\mathrm{Y}=\mathrm{y})}$$

代表的な確率分布としては、次のようなものがあげられる。

- ベルヌーイ分布 ： コイントス（裏と表）のイメージ
- マルチヌーイ（カテゴリカル）分布 ： さいころを転がすイメージ
- 二項分布 ： ベルヌーイ分布の多試行版
- ガウス分布 ： 釣り鐘型の連続分布

第3章:情報理論

情報理論とは、情報とは何かを定義し、情報の取り扱いについて考察することを目的としている。この中で情報量を定義すること

事象 x がおきる確率を P(x)と表し、事象 x がおきたことを知った情報量を I(x) と表すと情報量は、以下の確率の関数となる。

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

このとき、情報量の単位は、対数の底が 2 の時はビット(bit)、e の時はナット(nat)となる

この情報量の期待値 E(I(x)) = H(x)は次のようにあらわされ、シャノンエントロピーという。

$$\begin{aligned} \mathrm{H(x)} &= \mathrm{E(I(x))} \\ &= \mathrm{-E(\log(P(x))} \\ &= -\Sigma(\mathrm{P(x)\log(P(x)))} \end{aligned}$$

同じ事象、確率変数における異なる確率分布 P, Q の違い(距離)をあらわす概念として、カルバック•ライブラー　ダイバージェンス(KL ダイバージェンス)を用い、以下で定義される。

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

また、　交差エントロピーも同様に 2 つの確率分布がどの程度離れているかを判定するために用いられ、下記のように定義される。機械学習では誤差関数として用いられる。

$$\begin{aligned} \mathrm{H(P, Q)} &= \mathrm{H(P)} + D_{\mathrm{KL}}(P||Q) \\ &= \mathbb{E}\mathrm{x{\sim}p} \log \mathrm{Q(x)} \end{aligned}$$