応用数学レポート

１.線形代数

スカラー：大きさのみを持つ量、ベクトルに対する係数となる．

**ベクトル：**大きさと向きを持つ．スカラーのセットで表示される．

　　　　　複数の要素を持てるため，物の属性をベクトル的に表せる．

**行列：**スカラーを表もしくはベクトルを並べたもの．ベクトルの変換に用いる装置と見なせる．

**行列のベクトルへの作用**

n次元ベクトルにn次正方行列を作用させると新たなn次元ベクトル．

n次元ベクトルにm×n型の行列を（左から）作用させるとm次元ベクトル

**単位行列：**対角成分が1でそれ以外の成分が0の行列．

**逆行列：**行列の逆数みたいなもの．AA−1=A−1A=I

　　　　　掃き出し法によって、逆行列を求めることが出来る

**固有値と固有ベクトル**

ある正方行列 A に対して，

Ax→=λx →の式が成り立つような特殊なベクトル x→ と右辺の係数λ が存在し、それぞれ行列Aに対する，固有ベクトル，固有値という．

### **固有値 λ の求め方：**Ax→x→≠0→ より，(A−λI) に逆行列は存在しないので，|A−λI|=0　から求められる．

### **固有ベクトル x の求め方：**求めた λ をAx→=λx→ に代入して求める．

# **固有値分解**

# n次正方行列 A が固有値 λ1,λ2,… と固有ベクトル x1→,x2→,… を持つとする．

# この固有値を対角線上に並べた行列（非対角成分は

# Λ=(λ1　0⋯0

# 0λ2⋯0

# ⋮⋮⋱⋮

# 0⋯0λn)と，

# それに対応する固有ベクトルを並べた行でA=VΛV−1　と変形できる．このように正方行列 A を上の３つの行列積に分解することを固有値分解という．

### **固有値分解の利点**

* 行列の累乗の計算が容易になる．
* Λ から行列の性質を判断できる
* Λ は対角行列なので，小さな値を 0 とすることが可能．計算量を落とせる．

# **特異値分解**

　固有値分解は正方行列でしかできないが，非正方行列でも似たようなことはできる．

非正方行列M を特異値分解する．

Mv→=σu→、MTu→=σv→

このような特殊な単位ベクトルがあるならば，特異値分解できる．

M=UΣV−1

U は u→ を並べたもの，V は v→ を並べたもの．

　　　u→ を左特異ベクトル，　v→ を右特異ベクトルという．

## **特異値の求め方**

MV=UΣ、MTU=VΣTより

M=UΣV−1MT=VΣTU−1

これらの積は

MMT=UΣV−1VΣTU−1=UΣΣTU−1

つまり， MMT を固有値分解すれば，その左特異ベクトルと特異値の２乗が求められる．（ただし，左特異ベクトルは単位ベクトルである）

同様に，MTM=VΣU−1UΣTV−1=VΣTΣV−1

から，右特異ベクトルと特異値の2乗が求められる．

### **特異値分解の利点**

非正方行列は累乗できない．

Σ から行列の性質を判断できる

Σ は対角行列なので，小さな値を 0 とすることが可能．計算量を落とせる．（特異値は降順に並べる．）

* 特に、機械学習に関係する事としては以下の例がある。 与えられた２つの画像の特異値の大きい部分が同じであれば、それらは同じではないか。 機械学習の前処理において、特異値分解してあるもの同士でまとめると、教師なしでもそれらは同じものだと言えるのではないか。

２.確率・統計

集合とは、事象の集まり

確率は、その事業eventの発生する頻度、頻度確率（客観確率）またはその信念の度合い（主観確率＝ベイズ確率）

統計は記述統計と推測統計に分けられ、

記述統計は集団の性質を要約し記述すること、

推測統計は集団から一部を取り出し(標本）て元の集団（母集団）の性質を推測するもの

集団の性質を表すために確率分布を利用することができ、期待値や分散／標準偏差、分布間の相関を表す 共分散などが指標として用いられる。

条件付き確率：P（B｜A)=P(A交わりB）／P（B）

ベイズの法則：P（A)P(B|A)=P(B)P(A|B)

確率変数：事象と結び付けられた数値、

確率分布：事象の発生する確率の分布

期待値：分布における確率変数の平均の値、またはありえそうな値（離散的な場合と連続的な場合）

分散：データの散らばり具合＝二乗の平均ー平均の二乗　（標準偏差は分散の平方根）

様々な確率分布

ベルヌーイ分布：コイントスのイメージ

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

二項分布：ベルヌーイ分布の多試行版

　期待値　E(x)=np 分散V(x)＝np(1-p)

ガウス分布（正規分布）：釣鐘型の連続分布

３.情報理論

情報の変化は直感的には比率で捉えていると考えられる。珍しい（発生確率が低い）事象の方が情報量は大きいと考えられ、その大きさは

自己情報量　I(x)=ーlog2(P(x))=log(W(x))と表される。

on/offのスイッチで情報を伝える時に、情報の種類数に対して必要なスイッチの数は、事象の数Wのlogを取ることで求められる

概念的に、確率が低い事象程、自己情報が大きくなる。情報エントロピーとはわからなさ　あるいは不確実性

わからなさは　新たに何か分かった時の平均情報量と等しい

情報エントロピー＝平均情報量

●シャノンエントロピーとは、自己情報量の期待値で、

H（x）＝　E(I(x))= ーE(log(P(x))=ーΣ(P(x)log(P(x))

一般的には情報量が最大になる値を考えることが多い。機械学習では誤差関数代わりに使用できる。

相対エントロピーは別名カルバックライブラーの情報量と呼ばれる。確率分布が異なっていれば、情報量があると見做す。

●カルバック　ライブラー　ダイバージェンスとは、同じ事象・確率変数における　異なる確率分布の違いを表したもの。情報利得。自己情報量同士の差分に対する平均（期待値）を意味し、異なる確率分布間の違いを表す。（例：普通のコインと不正なコインの、表と裏の出る確率の違い）

例えば、とある低い確率で起きる事象があった場合の分布と、実際に起きた場合の分布の違いなど。

●交差エントロピー：Q（想定）についての自己情報量をP（現実）の分布で平均を取っている。KLカルバック　ライブラー　ダイバージェンスの一部を取り出す 。Q（想定していた信号）についての自己情報量をP（現実の信号）の分布で平均

結合エントロピー　H(A,B)=H(A)+H(B|A)=H(B)+H(A|B)

H(B|A)はAに関する情報を抜いたBから得られる平均情報量、条件付きエントロピー

相互情報量とは、不確実性（情報エントロピー）の減少量とみなせる。

I（B;C)＝元々のBの不確実性ーCが分かった後のBの不確実性

　　　＝H(B)ーH(B｜C)