応用数学

線形代数

スカラーとベクトルの違い
・スカラー：普通の数。ベクトルの係数になれる。
・ベクトル：大きさと向きを持ち、矢印で図示される。スカラーのセットで表される。

行列
・ベクトルの変換に使う。（行列とベクトルの積）
　変換後の要素のあるひとつは、元の要素のすべてから影響を受けている。
・行列との積を取ることで、ベクトルの要素数の変更ができる。

固有値と固有ベクトル
・行列とベクトルの積が、スカラーとベクトルの積で表されることがある。
　$Ax=\lambda x$
　この時、ベクトル（ｘ）を固有ベクトル、スカラー（λ）を固有値と呼ぶ。
　固有ベクトルで表すことで、行列とベクトルの積の計算を簡易に表せるメリットがある。
・固有値=単位行列×固有値で表される。
　·行列式の大きさ
　$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$

固有値分解
・固有値を持つ正方行列が、固有ベクトル、固有値、固有ベクトルの逆行列の合わせ技で表される。
・固有値分解を行うことで、行列の特徴を表すことができる。
・固有値分解の手順
　（１）固有値を求める。
　　$A\vec{x} = \lambda vec x$
　　$\rightarrow (A - \lambda I) = 0$
　　$\rightarrow x \neq 0$ より $|A - \lambda I| = 0$ を解く
　（２）固有ベクトルを求める。
　（３）固有ベクトルの逆行列を求める。
　　３次行列の場合は掃き出し法を使う。

特異値分解
・正方行列ではない行列について、固有値分解と似たようなことができる。
　以下が成り立つ場合、特異値分解ができる。

　$Mv=\sigma u$
　$M^T U=\sigma v$
　$M=USV^{-1}$

　Ｕ：行列ｕを展開したもの。
　V^{-1}：行列ｖを展開し、逆行列を取ったもの。
　Ｕ、Ｖの大きさは１に調整する。
　Ｓ：固有値を対角線上に並べた行列

・特異値ベクトル：大きさを１に調整されている。
・特異値は、MM^T、M^TM'（正方行列となる）を固有値分解して得られた特異値の２乗の平方根を取って求める。

確率・統計

確率

頻度確率（客観確率）とベイズ確率（主観確率）
・頻度確率（客観確率）　：実験して客観的に調べた確率
・ベイズ確率（主観確率）：信念の度合いを数字で表したもの

独立な事象の同時確率
・それぞれの事象が発生する確率の積で表される。
・Ｐ（Ｘ＝ｘ、Ｙ＝ｙ）＝Ｐ（Ｘ＝ｘ）Ｐ（Ｙ＝ｙ）＝Ｐ（Ｙ＝ｙ、Ｘ＝ｘ）

条件付確率
・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）＝Ｐ（Ｙ＝ｙ、Ｘ＝ｘ）／Ｐ（Ｘ＝ｘ）*

ベイズ則
・Ｐ（飴玉）＝１／４
　Ｐ（笑顔｜飴玉）＝１／２
　Ｐ（笑顔）＝１／３
　Ｐ（飴玉｜笑顔）Ｐ（笑顔）＝Ｐ（笑顔｜飴玉）Ｐ（飴玉）
　Ｐ（飴玉｜笑顔）＝１／２×１／４÷１／３＝３／８
・ベイズ則は主観確率でも客観確率でも両方使える。

確率変数と確率分布
・確率変数：事象と結びつけられた数値。事象そのものと解釈することもできる。
・確率分布：事象の発生する確率の分布。

期待値
・期待値 $E(f)=\sum_{K=1}^{n} P(X=x_k)f(X=x_k)$
・連続値の場合
期待値 $E(f)=\int P(X=x_k)f(X=x_k)dx$
・期待値は平均値。

分散と共分散
・分散：データの散らばり具合。期待値からどれだけずれているか。
分散 $Var(f)$
$=E((f(X=x)-E(f))^2)$
$=E((f^2_{(X=x)})-E(f)^2$
・共分散：２つのデータ系列の傾向の違い。正…似た傾向、負…逆の傾向、ゼロ…関係性に乏しい。
共分散 $Cov(f、g)$
$=E((f(X=x)-E(f))(g(Y=y)-E(g)))$
$=E(fg)-E(f)E(g)$
・データ同士の関係性を考察する場合に活用する。

様々な確率分布
・ベルヌーイ分布：コイントスのイメージ。０か１か。
$P(x|\mu)=\mu^x(1-\mu)^{1-x}$
・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。
$P(x=1)=\lambda_1$
$P(x=2)=\lambda_2$
$P(x=3)=\lambda_3$
$\lambda_1+\lambda_2+\lambda_3=1$
・二項分布：ベルヌーイ分布の多試行版。
$P(x|\lambda、n)=(n!/x!(n-x)!)\mu^x(1-\mu)^{1-x}$
・ガウス分布：釣り鐘型の連続分布
$N(x;\mu、\sigma^2)=\sqrt{\frac{1}{2\pi\sigma^2}}\exp\left(\frac{-1}{2\sigma^2}\right)(x-\mu)^2$

情報理論

自己情報量
・情報の価値（珍しさ）を表す。珍しければ価値がある。→対数を取って表す。
・$I(x) = -\log(P(x)) = \log(W(x))$
　対数の底が２のとき，単位はｂｉｔ。
　対数の底がｅのとき，単位はｎａｔ。
・$A \times B = 2^{a} \times 2^{b} = 2^{a+b}$
　$I(x) = \log_{2} A \times B = a + b = \log A + \log B$
・情報量の定義
$$dI = \frac{dW}{W}$$
$$\int dI = \int \frac{dW}{W}$$
$$I = \int \frac{1}{W} dW = \log_{e} W$$
・人間はｌｏｇの見方でものを見ている。（情報量の増え方が多いと増えたことに簡単に気づける）

シャノンエントロピー
・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
・２つの事象の情報量（＝確率分布）の違いを表す。

交差エントロピー
・ＫＬダイバージェンスの一部を切り出したもの。
・Ｐを基準にＱがどれだけ異なるかを表す。