応用数学 第3章：情報理論

はじめに

「応用数学 第3章：情報理論」について、動画を見ながら公式を記載したり例題を解いたりしながら理解を深める。

自己情報量

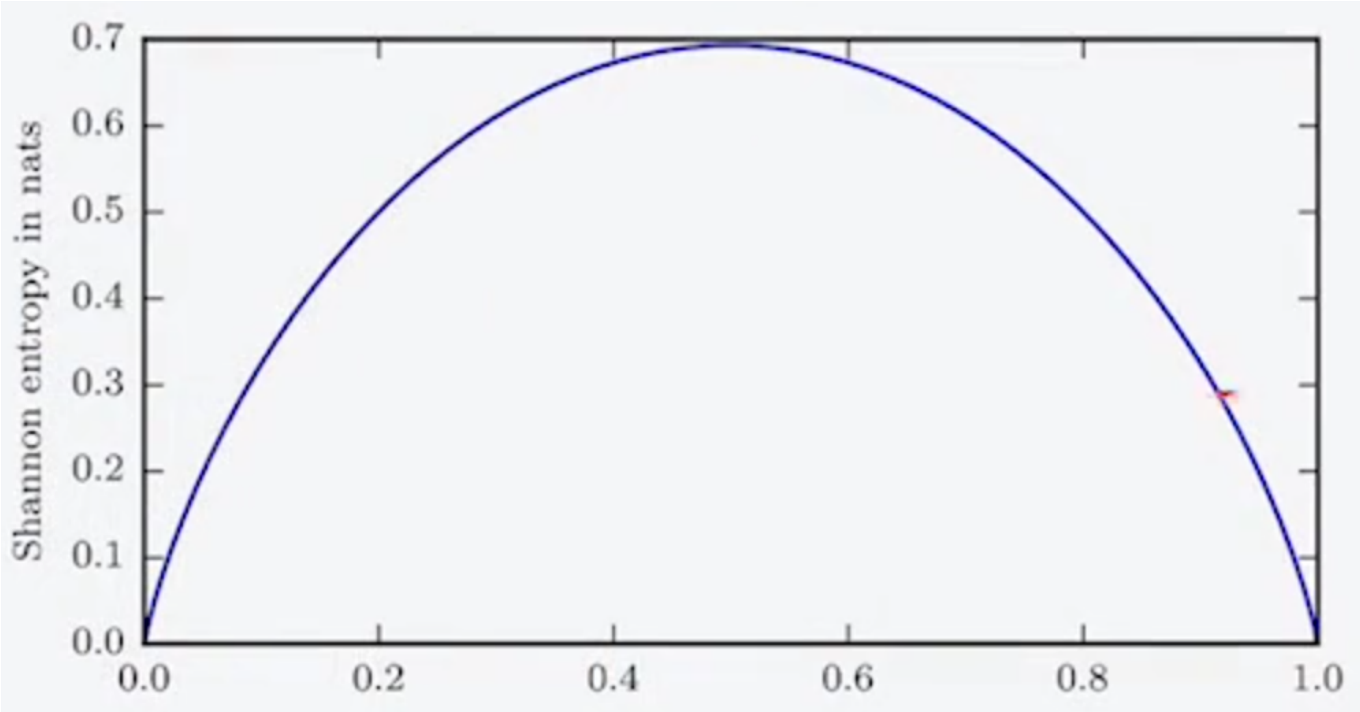
- 対数の底が2のとき、単位はビット（bit）
- 対数の底がeのとき、単位はナット（nat）
- PとWは逆数の関係

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

シャノンエントロピ

- コイントス、コインを投げた時にどれだけ新しく情報を得られるかを表したもの
- 横軸0.5、つまり偏りがないときは一番情報量が多い
- 横軸0.0、1.0、つまり必ず裏が出る、表が出ると決まっている場合は得られる情報なんてほとんどない
- 自己情報量の期待値
- 誤差関数として使うケースもある

$$\begin{aligned}
&H(x) \\
&= E(l(x)) \\
&= -E(\log(P(x))) \\
&= -\sum(P(x)\log(P(x)))
\end{aligned}$$



カルバック・ライブラー　ダイバージェンス

- 距離のようなもの
- 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
- Pが「最初に見積もった情報の珍しさ」、Qが「実際に測定してみて明らかになった情報の珍しさ」

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

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

$$E(f(x)) = \sum_x P(x) f(x)$$

$$D_{KL}(P||Q) = \sum_x P(x)(-\log(Q(x)) - (-\log(P(x))) = \sum_x P(x)\log\frac{P(x)}{Q(x)}$$

- シャノンエントロピーと似ているぞ･･･誤差関数にも使えるのでは･･･

交差エントロピー

- KLダイバージェンスの一部を取り出したもの
- Qについての自己情報量をPの分布で平均している

$$D_{KL}(P||Q) = \sum_x P(x)(-\log(Q(x))) - (-\log(P(x)))$$

$$H(P,Q) = H(P) + D_{KL}(P||Q)$$
$$H(P,Q) = -E_{x \sim P}\log Q(x) = -\sum_x P(x)\log Q(x)$$

演習問題

第 4 章　情報量

4.1

減多に起こらない事象が発生したという情報（例：伝説の生物ツチノコを発見したという情報）の重要度と，しばしば起こる事象が発生したという情報（例：残念なことに風呂場で G から始まる名前を持つ昆虫を発見してしまったという情報）の重要度とでは，どちらの情報により重要さがあるだろうか。これは直観的に前者であるとわかると思う。情報理論の世界の考え方もこの直観には従う。たとえば，10 本に 1 本当たりが含まれているクジに当選したという情報より，1000 本に 1 本当たりが含まれるクジに当選したという情報の方が，（つまり発生する確率の小さい事象の発生する情報の方が）より情報の量が多いと考えるのだ。具体的には 2 回に 1 回は起こるような事象（確率 1/2 の事象）が発生したという情報の大きさを 1bit の「（自己）情報量」として定義している。この情報量は足し合わせ可能な量だと考え，2 回に 1 回は起こるような事象 A と，また別の 2 回に 1 回は起こるような事象 B が同時に発生するという事象（確率 1/4 の事象）の情報量を 2bit，2 回に 1 回は起こるような事象 A と，また別の 2 回に 1 回は起こるような事象 B と，さらに別の 2 回に 1 回は起こるような事象 C が同時に発生するという事象（確率 1/8 の事象）の情報量を 3bit のように考えていく。この関係を数式で表すと以下のようになる。

$$\mathrm{I} = \log_2\left(\frac{1}{\mathrm{p}(X)}\right) = -\log_2\bigl(\mathrm{p}(X)\bigr)$$

※I は情報量， p (X)は事象 X の発生する確率。

この定義にしたがって，次の問いに答えよ。

問 4.1.1　1 枚のコインを 1 回投げて表が出たという事象の情報量は何 bit か。

問 4.1.2　2 枚のコインを 1 回投げてすべて表が出たという事象の情報量は何 bit か。

問 4.1.3　n 枚のコインを 1 回投げて 1 枚の表が出たという事象の情報量は何 bit か。

問4.1.1

- 回答

$$I = -\log_2\left(\frac{1}{2}\right) = -\log_2 2^{-1} = 1(bit)$$

- 考察
  - 1枚のコインを1回投げて表が出る確率P(X)=1/2

問4.1.2

- 回答

$$I = -\log_2\left(\frac{1}{4}\right) = -\log_2 2^{-2} = 2(bit)$$

- 考察
  - 2枚のコインを1回投げてすべて表が出る確率P(X)=1/2 x 1/2 = 1/4

問4.1.3

- 回答（★ 誤り ★）

$$I = -\log_2\left(\frac{1}{2}\right)^n = -\log_2 2^{-n} = n(bit)$$

- 考察（★ 誤り ★）
  - n枚のコインを1回投げて1枚の表が出る確率は、$(\frac{1}{2})^1$ x $(\frac{1}{2})^{n-1}$ = $(\frac{1}{2})^n$

- 回答（正答）

$$I = -\log_2\left(n \times 2^{-n}\right) = -\log_2(n) - (-n)\log_2 2 = -\log_2(n) + n(bit)$$

- 考察（正答）
  - n枚のコインを1回投げて1枚の表が出る確率P(X)=${}_n\mathrm{C}_1\left(\frac{1}{2}\right)^1$ x $\left(\frac{1}{2}\right)^{n-1}$ = $n \times \left(\frac{1}{2}\right)^n$
  - n枚のうち1枚だけ表が出る組合せは${}_n\mathrm{C}_1\left(\frac{1}{2}\right)^1$ 、残りn-1枚がそれぞれ1/2の確率で裏表どちらかが出る

第 6 章　対数と乗算除算の関係

問 6.1

実数 X，A，B に対して $X = AB$ という関係が成り立っている。次の式の空欄に入れることのできる適切な選択肢を選べ。

$$\log(X) = \boxed{}$$

ア　$\log\left(\frac{A}{B}\right)$　　イ　$\log(A) - \log(B)$　　ウ　$\log(A) + \log(B)$　　エ　$A\log(B)$

問 6.2

実数 X，A，B に対して　$X = \frac{A}{B}$　という関係が成り立っている。次の式の空欄に入れることのできる適切な選択肢を選べ。

$$\log(X) = \boxed{}$$

ア　$\log(AB)$　　イ　$\log(A) - \log(B)$　　ウ　$\log(A) + \log(B)$　　エ　$A\log(B)$

問 6.3

実数X，x_1，x_2，x_3，x_4に対して　$X = x_1x_2x_3x_4$　という関係が成り立っている。次の式の空欄に入れることのできる適切な選択肢を選べ。

$$\log(X) = \boxed{}$$

ア　$\log(x_1)\log(x_2)\log(x_3)\log(x_4)$　　イ　$\log(x_1x_2) - \log(x_3x_4)$

ウ　$\sum_{k=1}^{4}\log(x_k)$　　エ　$\prod_{k=1}^{4}\log(x_k)$

問6.1

- 回答
  - ウ

- 考察
    - 公式通り

問6.2

- 回答
    - イ

- 考察
    - 公式通り

問6.3

- 回答
    - ウ

- 考察
    - 公式通り

おわりに

本レポートでは、動画の要点を記載しながら最後に付属の演習問題を回答した。
動画では例題が少なかった。
自己情報量については演習問題で多少イメージはできたが機械学習にどう用いるのかまではイメージできていない。
シャノンエントロピーやカルバック・ライブラー ダイバージェンス、交差エントロピーについても同様なので、
確認テストやインターネットの記事を通してさらに理解を深めていこうと思う。