情報理論05

藤田一寿
津山工業高等専門学校情報工学科講師
電気通信大学先進理工学科協力研究員

ver. 20160512

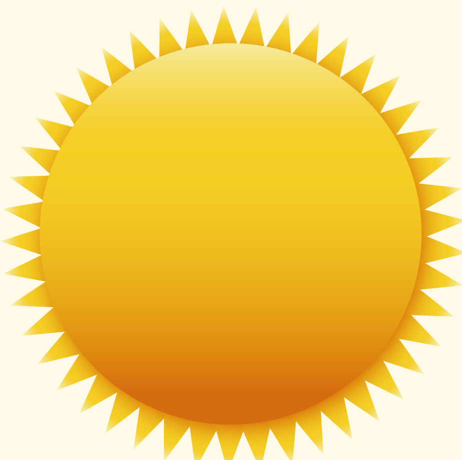
エントロピーの続き

面倒なのでlog2をlogと書きます。

天気の例

ある地区で各天気が起こる確率

晴れ



40%

曇り

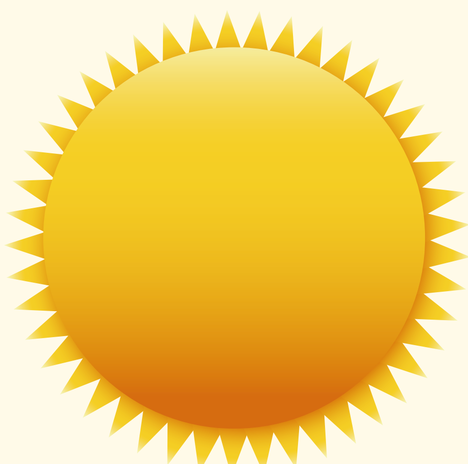
40%

雨

20%

天気予報で晴れと聞いたら

天気予報が晴れだった場合、晴れる可能性が高い。

晴れ	曇り	雨
		
80%	15%	5%

天気予報を聞く前のエントロピー

$$\begin{aligned} H &= -0.4\log 0.4 - 0.4\log 0.4 - 0.2\log 0.2 \\ &= 1.52 \end{aligned}$$

天気予報を知ったあとのエントロピー

$$\begin{aligned} H' &= -0.8\log 0.8 - 0.15\log 0.15 - 0.05\log 0.05 \\ &= 0.88 \end{aligned}$$

エントロピーが減ったということは、天気予報から情報を得て不確定さが減ったというように解釈できる。
天気予報から得た情報量は

$$I = H - H' = 0.64\mathrm{bit}$$

結合エントロピー

天気を例にもう少し考えてみる

- 明日の天気といえば、天気と気温がセットになっている。
- 天気を表す事象系をA
- 気温を表す事象系をB
- 天気と気温の複合事象(A, B)のエントロピーは

$$H(A, B) = -\sum_{ij} p(A_i, B_j) \log p(A_i, B_j)$$

結合エントロピー

熱と風邪の例

$$A : \begin{cases} A_1 & : \text{熱がある} \\ A_2 & : \text{熱がない} \end{cases} \qquad B : \begin{cases} B_1 & : \text{風邪を引いている} \\ B_2 & : \text{風邪ではない} \end{cases}$$

	B1風邪	B2風邪なし	p(Ai)
A1熱あり	0.55	0.05	0.6
A2熱なし	0.10	0.30	0.40
p(Bi)	0.65	0.35	

ここの確率は根拠があるわけではありません。

熱と風邪の結合エントロピー

$$H(A,B) = -0.55\log 0.55 - 0.05\log 0.05$$
$$- 0.1\log 0.1 - 0.3\log 0.3$$
$$= 1.54$$

結合エントロピーの方が、単一事象のエントロピーより大きい

熱のあるなしのエントロピー

$$H(A) = -0.6\log 0.6 - 0.4\log 0.4$$
$$= 0.97$$

風邪をひいたかひかないかのエントロピー

$$H(B) = -0.65\log 0.65 - 0.35\log 0.35$$
$$= 0.93$$

条件付きエントロピー

- AとBが独立でない場合、Aが分かっていた状態でのBのエントロピーを定義できる。
- 先ほどの風邪の例では、熱があれば風邪である可能性が高い。よって不確定さが減り、熱という情報を得たあとのエントロピーは結合エントロピーより小さいはずである。

条件付きエントロピー

A_iという事象が起こった状態でのBのエントロピーは

$$H(B|A_i) = -\sum_j p(B_j|A_i)\log p(B_j|A_i)$$

である。さらに、これのAについての期待値を求めると

$$\begin{aligned} H(B|A) &= \sum_i p(A_i)H(B|A_i) \\ &= -\sum_i p(A_i)\sum_j p(B_j|A_i)\log p(B_j|A_i) \\ &= -\sum_{ij} p(A_i,B_j)\log p(B_j|A_i) \end{aligned}$$

風邪と熱の例

$p(B_1|A_1) =$

$=$

$p(B_2|A_1) =$

$p(B_1|A_2) =$

$p(B_2|A_2) =$

	B_1風邪	B_2風邪なし	$p(A_i)$
A_1熱あり	0.55	0.05	0.60
A_2熱なし	0.10	0.30	0.40
$p(B_i)$	0.65	0.35	

風邪と熱の例

$$p(B_1|A_1) = \frac{p(A_1, B_1)}{p(A_1)}$$
$$= \frac{0.55}{0.6} = 0.92$$

$$p(B_2|A_1) = \frac{0.05}{0.6} = 0.08$$

$$p(B_1|A_2) = \frac{0.1}{0.4} = 0.2$$

$$p(B_2|A_2) = \frac{0.3}{0.4} = 0.75$$

	B_1風邪	B_2風邪なし	$p(A_i)$
A_1熱あり	0.55	0.05	0.60
A_2熱なし	0.10	0.30	0.40
$p(B_i)$	0.65	0.35	

熱があるときの風邪かどうかの条件付きエントロピー

$$H(B|A) = -\sum_i \sum_j p(A_i)p(B_j|A_i)\log p(B_j|A_i)$$

熱があるかどうか分かることによって得られた情報量は

$$\begin{aligned} I &= H(B) - H(B|A) \\ &= 0.36 \end{aligned}$$

熱があるときの風邪かどうかの条件付きエントロピー

$$\begin{aligned} H(B|A) &= -\sum_i \sum_j p(A_i) p(B_j|A_i) \log p(B_j|A_i) \\ &= -0.55 \times \log 0.92 - 0.05 \times \log 0.08 \\ &- 0.1 \times \log 0.25 - 0.30 \times \log 0.75 \\ &= 0.57 \end{aligned}$$

熱があるかどうか分かることによって得られた情報量は

$$\begin{aligned} I &= H(B) - H(B|A) \\ &= 0.36 \end{aligned}$$

エントロピーの性質

性質１

$$H(A,B) = H(A) + H(B|A)$$
$$= H(B) + H(A|B)$$

$$H(B|A) = H(A,B) - H(A)$$

演習

この性質を証明せよ

性質２

$$H(B|A) \geq 0$$

エントロピーは非負なので当然

性質3

$$H(A) + H(B) \geq H(A, B)$$

証明

$$\begin{aligned}
H(A) + H(B) &= -\sum_i p(A_i) \log p(A_i) - \sum_i p(B_j) \log p(B_j) \\
&= -\sum_{ij} p(A_i, B_j) \log p(A_i) - \sum_{ij} p(A_i, B_j) \log p(B_j) \\
&= -\sum_{ij} p(A_i, B_j)(\log p(A_i) + \log p(B_j)) \\
&= -\sum_{ij} p(A_i, B_j)(\log p(A_i) p(B_j))
\end{aligned}$$

$$H(A, B) = -\sum_{ij} p(A_i, B_j) \log p(A_i, B_j)$$

$$-\sum_{ij} p(A_i, B_j) \log p(A_i)p(B_j) \geq -\sum_{ij} p(A_i, B_j) \log p(A_i, B_j)$$

上記の式を簡単にすると

$$-\sum_i p_i \log q_i \geq -\sum_i p_i \log p_i$$

$$\sum_i p_i \log p_i - \sum_i p_i \log q_i \geq 0$$

$$\sum_i p_i \log p_i/q_i \geq 0$$

となり、これが成り立てば良い。

ここで、ギブスの不等式より

$$x - 1 \geq \ln x$$

$$-\ln x \geq 1 - x$$

$$\sum_i p_i \ln p_i / q_i \geq \sum_i p_i (1 - q_i / p_i)$$
$$= \sum_i (p_i - q_i) = 1 - 1 = 0$$

logはlnの正の定数倍なので

$$-\sum_{ij} p(A_i, B_j) \log p(A_i) p(B_j) \geq -\sum_{ij} p(A_i, B_j) \log p(A_i, B_j)$$

が成り立つ。

性質４

$$H(A) \geq H(A|B), H(B) \geq H(B|A)$$

性質１，３を使えば簡単に証明できる。

性質５

$$H(A, B) \geq H(A), H(B)$$

性質１とエントロピーの非負性で証明できる。

演習

- 2つのサイコロをランダムに振った時，一方のサイコロの目の確率変数をX，他方のサイコロの目の確率変数をYとする．ただし，すべての目は等確率に出現するとする．
  - H(X)を求めよ．
  - H(X+Y)を求めよ．
  - H(X|X+Y)を求めよ．
  - I(X, X+Y)を求めよ．

次回は相互情報量、KLダイバージェンス