第三章　情報理論

1) 自己情報量・シャノンエントロピーの定義を確認する。
2) KLダイバージェンス・交差エントロピーの概要を知る。

自己情報量（選択情報量）

事象Eが起こる確率をP(E)とするとき、事象 E が起こったことを知らされたとき受け取る（選択）情報量I(E) を

$$I(E) = \log \frac{1}{P(E)} = -\log P(E)$$

と定義する。
起こりにくい事象（＝生起確率が低い事象）の情報量ほど、値が大きい。

対数の底が2のとき，単位はビット(bit)
対数の底がネイピアのeのとき，単位は(nat)　　　　※natualの略

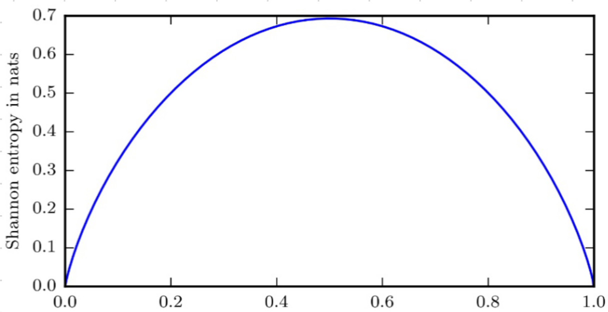
エントロピーの基本的性質 [編集]

1. 情報量は確率だけによって決まる。
2. 情報量は非負の値または無限大を取る。
3. nビットのビット列の空間（情報源）から（一様ランダムとは限らない方法で）ランダムにビット列を選んだときのエントロピーは、n以下になる。エントロピーがnになる必要十分条件は、ビット列が一様ランダムに選ばれることである。
4. 確率変数XとYが独立である必要十分条件は、$H(X)+H(Y)=H(X,Y)$が成立することである。

平均情報量（平均エントロピー/シャノンエントロピー/情報論のエントロピー）：自己情報量の期待値。

※微分エントロピーともいうが、微分してるわけではない。

$$H(x) = E\big(I(x)\big)$$
$$= -E\Big(\log\big(P(x)\big)\Big)$$
$$= -\sum\Big(P(x)\log\big(P(x)\big)\Big)$$



用途：例えば、「自己情報量の期待値が最大」となるようなxの値を求める。。など。

カルバック・ライブラー ダイバージェンス(カルバック・ライブラー 距離)

※厳密に数学的「距離」を満たしていないため、「ダイバージェンス」と呼ぶ
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。

離散確率分布PのQに対するカルバック・ライブラー情報量は以下の式で定義される。連続確率分布では積分する。
Qの自己情報量からPの自己情報量を引いて平均を取ったもので、分布間の距離のように考えることができる。非負の値を取る。

交差エントロピー

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

$$H(P,Q) = H(P) + D_{\mathrm{KL}}(P\|Q)$$

$$H(P,Q) = -\mathbb{E}_{\mathrm{x}\sim P}\log Q(x)$$