情報理論

1　エントロピー

定義 1.1. $(\Omega, \mathcal{A}, P)$ を確率空間、$X : \Omega \to \mathbb{R}$ を離散型確率変数とする。このとき、$X = x$ であるときの**自己情報量** $I(x)$ を

$$I(x) := -\log_2 P(x)$$

で定義し、X のエントロピー $H(X)$ を

$$H(X) := -\sum_{x \in X(\Omega)} P(x) \log_2 P(x)$$

で定義する。

例 1.1. あるスーパーで、１時間に８０人の客が来店した。来店した客の性別の内訳と、客がマイバッグを所持しているかの内訳は以下の通りであった。

(1) 性別が男でマイバッグを持っていた客の数 : 10
(2) 性別が女でマイバッグを持っていた客の数 : 10
(3) 性別が男でマイバッグを持っていなかった客の数 : 40
(4) 性別が女でマイバッグを持っていなかった客の数 : 20

このとき、客がマイバッグを所持しているかを知った時のエントロピーを求める。男性を M、女性を F で表し、マイバッグを持っていたということを 1、マイバッグを持っていなかったということを 0 で表すことにする。標本空間 Ω を $\Omega = \{(M,0),(M,1),(F,0),(F,1)\}$ で定義する。事象空間 $\mathcal{A}$ を $\mathcal{A} = \Omega$ で定義し、確率 $P : \mathcal{A} \to \mathbb{R}$ を $P((M,0)) = 1/2, P((M,1)) = 1/8, P((F,0)) = 1/4, P((F,1)) = 1/8$ で定義する。$(\Omega, \mathcal{A}, P)$ は確率空間である。確率変数 $X : \Omega \to \mathbb{R}$ を $X((x,y)) = y$ で定義する。$X = 0$ であるときの自己情報量は $I(0) = -\log_2 3/4$ であり、$X = 1$ であるときの自己情報量は $I(1) = -\log_2 1/4$ である。よって、$H(X) = -(3/4)\log_2 3/4 - (1/4)\log_2 1/4$ である。

2　カルバック・ライブラーダイバージェンス

定義 2.1. $(\Omega, \mathcal{A}, P), (\Omega, \mathcal{A}, Q)$ を確率空間、$X : \Omega \to \mathbb{R}$ を離散型確率変数とする。このとき、P, Q の**カルバック・ライブラーダイバージェンス** $D_{KL}(P\|Q)$ を

$$D_{KL}(P\|Q) := \sum_{x \in X(\Omega)} P(x)(\log_2 P(x) - \log_2 Q(x))$$

で定義し、P, Q の交差エントロピー $H(P, Q)$ を

$$H(P, Q) := -\sum_{x \in X(\Omega)} P(x) \log_2 Q(x)$$

で定義する。以下のことに注意せよ。

(1) $D_{KL}(P \| Q) = 0$ となるのは $P = Q$ であるときに限る。
(2) $D_{KL}(P \| Q), D_{KL}(Q \| P)$ は一般に等しくない。

参考文献

[1] 小縣信也., 斎藤翔汰., 溝口聡., 若杉一幸. (2021). ディープラーニング E 資格エンジニア問題集 インプレス.

[2] 瀧雅人. (2017). これならわかる深層学習 講談社