

MA17_情報理論1-1

情報量とは？

情報の量。

珍しい情報ほど価値がある。

珍しさの測り方は、何回に一回出るか？などと確率で表現する。

自己情報量

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

珍しいことが起こる(確率が小さい)と $I(x)$ は小さくなることを示している。

しかし、**珍しい情報ほど価値があるのが『情報量』**なのに値が小さくなっている。

だから、逆数で表した表現(Wで表す)で『情報量』を表すこともある。

それが下記の式。

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

なぜ、情報量はlogを用いて表現されるか？

人間は何か量が増えた時、足し算で増えると認識する。

しかし、同時確率等を考えると、事象の積であらわされてしまう。

これを**加算であらわす形に変換するためにlogを用いている**。

その他に、確率を用いているのでlogを使うことで珍しい事象(確率が小さい)を大きい値として表現できる。

底を2とした場合の単位は bit

底を e とした場合の単位は nat

MA18_情報理論1-2

人間は、どうやって情報量が増えたことを認識しているのか？

元々の事象の量に対して新たに増えた事象の量の比を、

どうやら増えた情報と認識しているようだ。

式で表すと下記である。

$$\Delta I = \frac{\Delta W}{W}$$

積分すると...

$$I = \log_e W \text{で表現される。}$$

このことから、人間は人間は情報の増え方を対数で見ていると考えられる。

シャノンエントロピー(平均情報量 or 自己情報量の期待値)

自己情報量の期待値として表現される。

$$H(x) = E(I(x))$$

$$= -E(\log(P(x))) = -\sum (P(x)\log(P(x)))$$

MA19_情報理論2

自己情報量により、ある事件の自体の情報量を見積れる。

シャノンエントロピーにより、そういった事件が様々な確率で出現されるときに平均的にどれくらい期待できるかを見積れるようになった。

確率分布が違った場合、それって情報量にどのくらい違いがあるのだろうか？

カルバック・ライブラー ダイバージェンス

2つの確率分布の違いによる情報量の大きさの違いを計算するために使用される。情報量の大きさは確率分布によって決まり、確率分布が違えば、全体の情報量の大きさが変わるということを示している式である。

式は下記である。

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

新しい確率分布 $P(x)$ を重みとして平均をとることがポイント。

※ P と Q を逆にすると値が異なるため、厳密には距離ではない。

MA19_情報理論3

交差エントロピー

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

$$H(P, Q) = - E_{x \sim P} \log Q(x)$$