２０１９年４月２２日（川端）

データ圧縮基礎資料

4　エントロピーおよび関連する情報量

4.1　エントロピー, ダイバージェンス, 相互情報量

本節以下では, $\mathcal{X}$ は離散集合とする．有限集合であればことわる．確率変数 (X,Y) のエントロピー $H((X,Y))$ を $H(X,Y)$ と略記し確率変数 X と Y の同時エントロピーと定義する．$H(X,Y) = H(Y,X)$ が成り立つ．$H(X,Y) - H(X)$ を $H(Y|X)$ と書いて, X に関する Y の条件付きエントロピーいう．$H(X) = -E\log p(X)$ および $H(X,Y) = -E\log p(X,Y)$ から $H(Y|X)$ に対する別表現 $H(Y|X) = -E\log p(Y|X)$ が得られる.

次に $I(X,Y)$ を $H(Y) - H(Y|X)$ と定義する．これを X と Y の相互情報量という．これは $I(X,Y) = H(Y) + H(X) - H(X,Y)$ とも書ける．即ち X と Y について対称な量である．また

$$I(X,Y) = E\log\frac{p(X,Y)}{p(X)p(Y)}$$

なる表現も得られる.

さて，$\mathcal{X}$ 上の確率分布 p と q について

$$D(p||q) := \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}$$

を q に関する p の Kullback-Leiber のダイバージェンス (KL 分離度) という.

これについて以下の定理が成り立つ.

定理 1

$$D(p||q) \geq 0,$$

ただし, 等号は $p = q$ のときのみ成立.

証明：

$$\begin{aligned} D(p||q) &= \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} \\ &\geq \sum_{x:p(x)>0} p(x)\frac{1}{\ln 2}(1-\frac{q(x)}{p(x)}) \\ &= \frac{1}{\ln 2}(\sum_{x:p(x)>0} p(x) - \sum_{x:p(x)>0} q(x)) \\ &\geq 0 \end{aligned}$$

等号成立条件は容易にチェックできる．証明終わり.

エントロピーは確率分布の上に凸な関数であるが KL 分離度はその接触変換[1] になっており，非負性はそのことからもいえる.

[1]一般に，上に凸な関数 $f:\mathcal{R}^k \to \mathcal{R}$ に対して，グラフ $y = f(\mathbf{x})$ を考えると，定義域における点 $\mathbf{x}_0$ における接触面の方程式は $y = f(\mathbf{x}_0) + (\mathbf{x}-\mathbf{x}_0)\cdot\nabla f(\mathbf{x}_0)$ となる．従って，後者引く前者 ($\mathbf{x}_0$ における $f(\mathbf{x})$ の接触変換) について不等式 $(\mathbf{x}-\mathbf{x}_0)\cdot\nabla f(\mathbf{x}_0) - (f(\mathbf{x}) - f(\mathbf{x}_0)) \geq 0$（，但し等号は $\mathbf{x} = \mathbf{x}_0$ のときのみ）が成り立つ．従って，$H(\mathbf{p})$ の $\mathbf{q}$ における接触変換は $\nabla H(\mathbf{q}) = \log(\mathbf{1}/\mathbf{q}) - (\log\mathbf{e})\mathbf{1}$ に注意すると $(\mathbf{p}-\mathbf{q})\cdot(\log\frac{\mathbf{1}}{\mathbf{q}} - (\log\mathbf{e})\mathbf{1}) + \mathbf{H}(\mathbf{q}) - \mathbf{H}(\mathbf{p}) =: \mathbf{D}(\mathbf{p}||\mathbf{q})$ となる.

分布空間を $\mathcal{X}\times\mathcal{Y}$ にもつ二つの確率分布 $p(x,y), q(x,y) := p(x)p(y)$ について，

$$\begin{aligned} D(p||q) &= \sum_{x,y\in\mathcal{X}\times\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= H(X)+H(Y)-H(X,Y) \\ &= I(X;Y) \end{aligned}$$

が成り立つ．従って，$I(X;Y)\geq 0$ が成り立つ．また等号成立条件は $p(x,y)=p(x)p(y)$，即ち X,Y が独立な確率変数のときそのときのみである．

$\mathcal{X}$ 上の $D(p||q)$ について，$q(x)=(1/\sharp\mathcal{X})$ と置くと，$D(p||q)=\log\sharp\mathcal{X}-H(X)$ となる．従って，$\log\sharp\mathcal{X}-H(X)\geq 0$ が成り立つ．さらに等号は $p(x)=(1/\sharp\mathcal{X})$ の場合だけである．

さて，

$$H(X|Y) = \sum_x p(x)\sum_y p(y|x)\log\frac{1}{p(y|x)}$$

であるが，内側の和はエントロピーであるから非負かつ上から $\log\sharp\mathcal{Y}$ で上界されている．従ってその平均も同様．まとめると

$$0 \leq H(X|Y)\leq\log\sharp\mathcal{Y}$$

4.2 条件付相互情報量

２つの確率変数 Y,Z の確率変数 X に関する条件付同時分布を考える．今 $X=x$ の条件で Y と Z は相互情報量を有するが，これを $I(Y;Z|X=x)$ と書くことにしよう．別な表現は

$$D(p_{YZ|X}(\cdot,\cdot|X=x)||p_{Y|X}(\cdot|X=x)p_{Z|X}(\cdot|X=x))$$

である．

定義 1

$$I(Y;Z|X)=\sum_x p_X(x)I(Y;Z|X=x)$$

を X に関する Y と Z の相互情報量という．

これは明らかに非負の量である．

定理 2

$$I(Y;Z|X)\geq 0 \tag{1}$$

次のような変形がある．

定理 3

$$\begin{aligned} I(Y;Z|X) &= E\log\frac{p(Y,Z|X)}{p(Y|X)p(Z|X)} \\ &= -H(Y,Z|X)+H(Y|X)+H(Z|X) \\ &= H(Y|X)-H(Y|Z,X) \\ &= H(Z|X)-H(Z|Y,X) \end{aligned}$$