パターン認識と機械学習

1.6 情報理論

05T4074A
久保田　敦

目的

- パターン認識や機械学習のテクニックをより発展させるのに有用な概念を学ぶ。

目次

- 情報量
- エントロピー（離散確率変数）
- logからlnへ
- エントロピー（連続変数）
- 同時分布

情報量

- 情報量h＝xの値を得たときの驚きの度合
- 単位：ビット

$$\boxed{h(x,y)=h(x)+h(y)} \quad \boxed{p(x,y)=p(x)p(y)}$$

$$h\big(p(x,y)\big)=h(p(x))+h(p(y))$$

$$h\big(p(x)p(y)\big)=h(p(x))+h(p(y))$$

$$\log_2 p(x)p(y)=\log_2 p(x)+\log_2 p(y)$$

$$\boxed{h(x)=-\log_2 p(x)}$$

エントロピー

- 別名：平均情報量

$$H[x] = -\sum_{x} p(x) \log_2 p(x)$$

確率　　　情報量

- 一様な分布よりも非一様分布の場合の方がエントロピーは小さい

エントロピー計算

- 一様な分布（8個の事象が同確立）

$$H[x] = -8 \times \frac{1}{8} \log_2 \frac{1}{8} = 3\text{ビット}$$

- 非一様な分布（ $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}$ ）

$$H[x] = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8}$$

$$-\frac{1}{16}\log_2\frac{1}{16} - \frac{4}{64}\log_2\frac{1}{64} = 2\text{ビット}$$

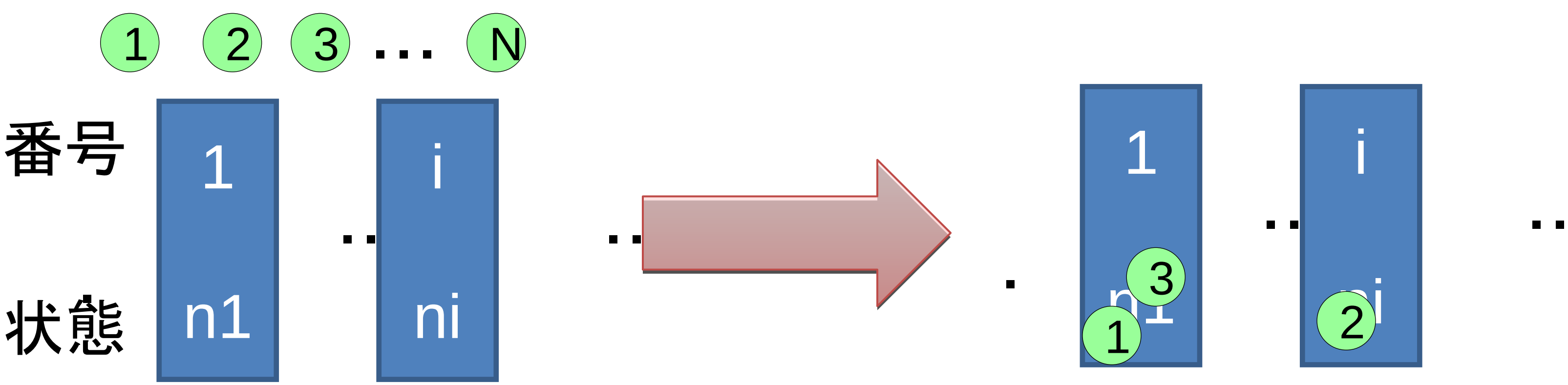
※符号化に用いられる

logからlnへ

- 関連性を考えて自然対数に
- 単位：ナット

別例：N個の物体を箱に振り分ける

- それぞれの箱の中で物体の入れた順番は区別しない



エントロピー（2）

- 物体の入れ方：多重度 $W = \dfrac{N!}{\prod_i n_i!}$
- 箱の中の物体の状態：ミクロ状態
- 物体の占有数：マクロ状態 $\dfrac{n_i}{N}$

箱＝X（離散確率変数）の状態xi
p(X=xi)=pi＝i番目の箱に物体が割り当てられる確率

$$H[p] = -\sum_i p(x_i) \ln p(x_i)$$

最大のエントロピー

$$H[p] = -\sum_i p(x_i)\ln p(x_i) + \lambda\left(\sum_i p(x_i) - 1\right)$$

エントロピーが最大になる分布：$p(x_i)$が等確率

$$p(x_i) = \frac{1}{M}$$

（M＝niの総数）

最大エントロピー

$$H = \ln M$$

連続変数

- 連続変数に対するエントロピー
（微分エントロピー）

$$H[X] = -\int p(X_i) \ln p(X_i) dx$$

X ：連続変数のベクトル

p(x) ：連続変数の分布

連続変数の最大エントロピー

エントロピーが最大になる分布：ガウス分布

$$p(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

ガウス分布の微分エントロピー

$$H[x] = \frac{1}{2}\left\{ 1 + \ln(2\pi\sigma^2) \right\}$$

同時分布

- 変数xとyから同時分布p(x,y)

  xが既知のとき必要な情報は $-\ln p(y|x)$

$$H[y|x] = -\iint p(y,x)\ln p(y|x)dydx$$

  xに対するyの条件付きエントロピー

- 確率の乗法定理から

$$H[y,x] = H[y|x] + H[x]$$