数字基础 8th

13 5
2 4 6

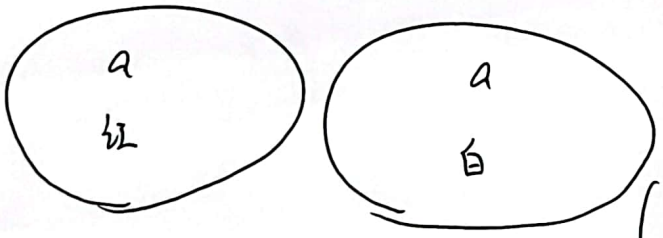
1、 $P(A) = \frac{1}{6} \times \frac{5}{6} \times 2 = \frac{5}{36} \times 2 = \frac{5}{18}$

$P(B) = \frac{1}{6} \times \frac{5}{6} \times 2 + \frac{1}{36} = \frac{5}{18} + \frac{1}{36} = \frac{11}{36}$

$P(C) =$ (奇+奇)，偶+偶 $= \frac{3\times3+3\times3}{6\times6} = \frac{18}{36} = \frac{1}{2}$

$I_A = -\log(\frac{5}{18})$ $I_B = -\log\frac{11}{36}$ $I_C = -\log\frac{1}{2}$

2、



从 2a 个球中取球

概率空间

$$\begin{pmatrix} X \\ P(X) \end{pmatrix} = \begin{pmatrix} X=x_1 \cdots X=x_i \cdots X=x_{1c} \\ P(x_1) \cdots P(x_i) \cdots P(x_{1c}) \end{pmatrix}$$

$x_i = 1$ / 0 → 第 i 次取出白球

↓

第 i 次取出红球

① 有放回情况下，每个 $P(x_i)$ 概率都相同，因为每次放回都是同样的5-堆球，且 $\sum P_i = 1$

② 无放回情况下，每次的情况都与前几次抽到什么球有关，所以 P_i 都各不相同，且 $\sum P_i = 1$

根据熵的极值性：离散信息源中各消息等概率出现时熵最大

∴ 有放回的熵更大

3、 $D_{KL}(p\|q) = E_{x\sim p}[\log p(x) - \log q(x)] = E_{x\sim p}\left[\log \frac{p(x)}{q(x)}\right]$

$E_{x\sim p}[f(x)]$ $f(x)$ 关于 $p(x)$ 的期望

$\therefore D_{KL}(p\|q) = \left[\log \frac{p(x)}{q(x)}\right.$ 关于 $p(x)$ 的期望$\left.\right]$

$H(p,q) = -E_{x\sim p}\log q(x)$

样本 $(x_1, x_2 \cdots x_i, \cdots x_n)$

类 $(y_1, y_2, \cdots y_i, \cdots y_{|C|})$

假设 x_i 属于标签 y (真实)

⇓

转换成 one-hot

$p_i = (0, 0, \cdots \underset{\text{第}i\text{个}}{1} \cdots 0)^T$

对于 x_i，给定 p_i，p_i 为第 i 个分量为 1 的 one-hot 向量

q_i 为预测的标签，$q_i = f(x_i, \theta)$

$\therefore D_{KL}(p\|q) = (p_i)^T\left(\log \frac{p_i}{q_i}\right) = p_i^T \log p_i - p_i^T \log q_i$

$H(p_i, q_i) = -p_i^T \log q_i$

真实值的向量

$\therefore$ $D_{KL}(p\|q)$ 和 $H(p_i, q_i)$ 是等价的

针对 q_i 最小化交叉熵等价于最小化 KL 散度，因为 q_i 不参与被省略的一项

$\therefore \arg\min_{\theta} D_{KL}(p_i\|q_i) = \arg\min_{\theta} H(p_i, q_i)$

即 $\theta = \hat{\theta}$ 时，可以同时使 KL 散度和交叉熵函数最小

习题4 $P(X_1, X_2 \cdots X_n) = P(X_1)P(X_2|X_1) \cdots P(X_n|X_{n-1})$

$I(X_1; X_2, \cdots X_n) = H(X_1) - H(X_1|X_2, \cdots X_n)$

$= H(X_1) - \left(H(X_1, X_2 \cdots X_n) - H(X_2, \cdots X_n)\right)$

⇓ 熵的链式法则

$= H(X_1) - \left(H(X_1) + H(X_2|X_1) + H(X_3|X_1X_2) + \cdots H(X_n|X_1 \cdots X_{n-1})\right) + \left(H(X_2) + H(X_3|X_2)\right.$

$\left. + H(X_4|X_2X_3) + \cdots H(X_n|X_2X_3 - X_{n-1})\right)$

$= H(X_1) - \underbrace{\left(H(X_1) + \sum_{i=2}^{n} H(X_i|X_{i-1})\right)}_{①} + H(X_2) + \sum_{i=3}^{n} H(X_i|X_{i-1})$

$= -\sum_{i=2}^{n} H(X_i|X_{i-1}) + H(X_2) + \sum_{i=3}^{n} H(X_i|X_{i-1})$

$= H(X_2) - H(X_2|X_1)$

$= I(X_2; X_1) = I(X_1; X_2)$

① for 马尔科夫

$P(X_1, X_2) = P(X_1) \cdot P(X_2|X_1)$ ✓

$P(X_1, X_2, X_3) = P(X_1) \cdot P(X_2|X_1) \cdot P(X_3|X_1X_2) = P(X_1)P(X_2|X_1)P(X_3|X_2)$

$P(X_3|X_1X_2) = P(X_3|X_2)$

$P(X_1, X_2, X_3, X_4) = P(X_1) \cdot P(X_2|X_1) \cdot P(X_3|X_1X_2) \cdot P(X_4|X_1X_2X_3) = P(X_1)P(X_2|X_1)P(X_3|X_2) \cdot P(X_4|X_3)$

$P(X_4|X_1X_2X_3) = P(X_4|X_3)$

$\vdots$

$\therefore$ for $\forall$ $P(X_i|\prod_{k=1}^{i-1} X_k) = P(X_i|X_{i-1})$ (在马尔科夫链中)

$H(X_3|X_1X_2) = E\left(\log \frac{1}{P(X_3|X_1X_2)}\right) = E\left(\log \frac{1}{P(X_3|X_2)}\right) = H(X_3|X_2)$

$H(X_4|X_1X_2X_3) = E\left(\log \frac{1}{P(X_4|X_1X_2X_3)}\right) = E\left(\log \frac{1}{P(X_4|X_3)}\right) = H(X_4|X_3)$

$\vdots$

$H(X_i|\prod_{k=1}^{i-1} X_k) = H(X_i|X_{i-1})$ (在马尔科夫链中)