第8次作业

习题1.解(A). $P_1 = \frac{1}{6}\times\frac{5}{6}\times 2 = \frac{5}{18}$　$H_1 = -\log P_1 = -\log\frac{5}{18}$

(B). $P_2 = 1-\frac{5}{6}\times\frac{5}{6} = \frac{11}{36}$　$H_2 = -\log P_2 = -\log\frac{11}{36}$

(C) $P_3 = \frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{2} = \frac{1}{2}$　$H_3 = -\log P_3 = -\log\frac{1}{2}$

习题2.解：首先，来分析有放回的情况，假定取到的前后顺序不同，则事件不同（例如"红,白"和"白,红"是两种不同的情况）

对于每种情况来说 ~~$P_k =$~~ $P_i = (\frac{1}{2})^k$　$i=1,2\cdots 2^k$

熵值　$H_1 = -\sum_{i=1}^{2^k} P_i \log P_i = -\sum_{i=1}^{2^k} (\frac{1}{2})^k \cdot \log(\frac{1}{2^k}) = -\sum_{i=1}^{N} \frac{1}{N}\log(\frac{1}{N})$，为极值.　$N = 2^k$

再来分析不放回的情况，无论 $k>a$ 还是 $k\le a$，取到的k个球，总情况数 $N_1 = \cancel{C_{2a}^{k}}\ 2^{\min\{a,k\}}$　对于任意每种情况 ~~(在取了一个球之后)~~

~~一定存在~~ 一定不存在 $P'_i = P_i$　($i=1,2\cdots 2^{\min\{a,k\}}$) 恒成立.

即 P' 取不到熵函数在最大值 P 时的情况，此时熵值一定比有放回时小.

故有放回时的熵更大

习题3.证明：*来一步步地拆解这个问题~.

假设样本数量为n. 样本 $x_1\cdots x_n$. 它们分别对应于 $y_1\cdots y_n$ (这是类别)

然后有一系列的预测 $y'_1,\cdots y'_n$

给定评估函数 $f: R^{2n}\to R^{n}$.　$f(Y) = t$.　$Y = [y\ \ y']$　$y = \begin{bmatrix} y_1\\ \vdots \\ y_n\end{bmatrix}$ $y' = \begin{bmatrix} y'_1\\ \vdots \\ y'_n\end{bmatrix}$

若 $y_i = y'_i$ 则 $t_i = 1$ 否则 $t_i = 0$.

** 这样复杂了，因为我无法表示概率，更无法表示熵函数

假设样本数量为n.　样本 $x_1\cdots x_n$.　它们对应的类别分别是 $y_1\cdots y_n$.

我们给的预测类别分别是 $z_1\cdots z_n$.

仅仅就第i个样本 ($i=1,\cdots n$)　对真实情况，P_i 表示其他元素都为0，只有第i个数为1的列向量　这是真实情况的概率分布

因为此时只有第i个元素为1与第i个类对应 ($P=1$)对应

$q_i = \cancel{f(K, z_i)}\ \cancel{f(K, x_i, y_i, z_i)}\ f(K, x_i)$ (K为超参数)　q_i 为预测的概率分布情况.

$D(P_i \| q_i) = P_i^T \log P_i - P_i^T \log q_i$

定义 $\log x = \begin{bmatrix}\log x_1\\ \vdots \\ \log x_n\end{bmatrix}$　$x\in R^n$.

而交叉熵为 $-P_i^T \log q_i$

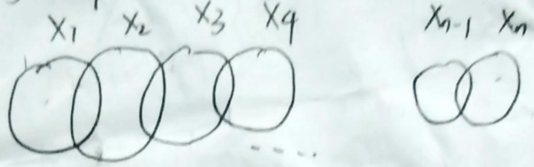
由于 $P_i^T \log P_i = 1\cdot\log 1 = 0$ 故KL散度为 $-P_i^T \log q_i$

两者取相同的最小值.(关于k的函数) 即交叉熵函数作为损失函数和用KL散度作为损失函数是等价的

题4. 解：由马尔科夫链的定义，由

$P(x_1, x_2 \cdots x_n) = P(x_1)\cdot P(x_2|x_1)\cdots P(x_n|x_{n-1})$ 可得它们的

venn图为



可以看到事件 x_i ($i=2\cdots n-1$) 只与 x_{i-1} 和 x_{i+1} 存在关联.

x_n 只与 x_{n-1} 存在关联.

$$I(x_1; x_2, \cdots x_n) = H(x_1) - H(x_1 | x_2, \cdots x_n)$$

$$= H(x_1) - \left(H(x_1, x_2 \cdots x_n) - H(x_2, \cdots x_n) \right)$$

$$= H(x_1) - \left(\sum_{i=1}^{n} H(x_i | x_{i-1}, \cdots x_1) - \sum_{i=2}^{n} H(x_i | x_{i-1}, \cdots x_1) \right)$$

其中，

$$① = \sum_{i=1}^{n} H(x_i | x_{i-1}, \cdots x_1) = H(x_1) + H(x_2|x_1) + \cdots + H(x_n | x_{n-1} \cdots x_1)$$

$$= H(x_1) + H(x_2|x_1) + H(x_3|x_2) + \cdots + H(x_n|x_{n-1})$$

$$② = \sum_{i=2}^{n} H(x_i | x_{i-1}, \cdots x_1) = H(x_2) + H(x_3|x_2) + \cdots + H(x_n | x_{n-1} \cdots x_2)$$

$$= H(x_2) + H(x_3|x_2) + H(x_4|x_3) + \cdots + H(x_n|x_{n-1})$$

$$① - ② = H(x_1) + H(x_2|x_1) - H(x_2)$$

原式 $= H(x_1) - (① - ②) = H(x_2) - H(x_2|x_1) = I(x_1; x_2)$