温兆和. 10205501432. 数学基础作业8.

1.解: $P(A)=2\times\frac{1}{6}\times\frac{5}{6}=\frac{5}{18}$.

故 $I(A)=\log\frac{18}{5}$

$P(B)=$ ~~$2\times\frac{1}{6}=\frac{1}{3}$~~ $1-P(\bar{B})=1-\frac{5}{6}\times\frac{5}{6}=\frac{11}{36}$

故 $I(B)=\log\frac{36}{11}$

在两个骰子中先确定其中一个骰子的点数再投另一个骰子. 无论第一个骰子是几, 点数之和都有一半概率是奇数, 一半概率是偶数. 故 $P(C)=\frac{1}{2}$

故 $I(C)=\log 2$

2.解: 假设取出的k个球中有x个红球.

则无放回时: $P(X=x)=\dfrac{C_a^x\,C_a^{k-x}}{C_{2a}^k}$

$$=\frac{\dfrac{a!}{x!(a-x)!}\cdot\dfrac{a!}{(k-x)!(a-k+x)!}}{\dfrac{(2a)!}{k!(2a-k)!}}$$

$$=C_k^x\,\frac{a!\,a!\,(2a-k)!}{(2a)!\,(a-x)!\,(a-k+x)!}$$

有放回时 $P(X=x)=C_k^x\left(\frac{1}{2}\right)^k$

故 $H(\text{有放回})=-\sum_{x=0}^{k}\left(\frac{1}{2}\right)^k C_k^x\left(\log\left(\frac{1}{2}\right)^k+\log C_k^x\right)$

$$>-\sum_{x=0}^{k}C_k^x\log C_k^x$$

由于 $\dfrac{a!\,a!\,(2a-k)!}{(2a)!\,(a-x)!\,(a-k+x)!}>1$ 且 $f(x)=-x\log x$ 递减

又 $H(\text{无放回})=$

$$-\sum_{x=0}^{k}\frac{a!a!(2a-k)!}{(2a)!(a-x)!(a-k+x)!}C_k^x\log\frac{a!a!(2a-k)!}{(2a)!(a-x)!(a-k+x)!}C_{k-x}.$$

故 H(有放回) $< -\sum C_k^x\log C_k^x < H$(无放回)

故有放回的情况熵更大

3. 证：在机器学习中，我们不妨用 $p(x)$ 表示真实的分类情况，而用 $q(x)$ 表示用模型预测的分类情况。损失函数是用来衡量真实值和预测值的距离的。在设计模型时我们要最小化这个损失函数。由于

KL 散度 $D_{KL}(p\|q)=E_{x\sim p}[\log p(x)-\log(q(x))]$

交叉熵 $H(p,q)=H(p)+D_{KL}(p\|q)$

由于 $H(p)$ 仅与 p 有关，与 p,q 间的差距无关

故 $\arg\min D_{KL}(p\|q)$ 与 $\arg\min H(p,q)$ 等价。

故用 KL 散度和用交叉熵作损失函数等价

4. 解：$I(X_1;X_2\ldots X_n)=H(X_1)-H(X_1|X_2\ldots X_n)$

$=H(X_1)+H(X_2\ldots X_n)-H(X_1\ldots X_n)$

其中，$H(X_1\ldots X_n)=H(X_1\ldots X_{n-1})+H(X_n|X_1\ldots X_{n-1})$

$=\cancel{H(X_1}\ldots\ldots=H(X_1)+H(X_2|X_1)+H(X_3|X_2,X_1)+\ldots$

$+H(X_n|X_1\ldots X_{n-1})$

由于在马尔可夫链中 $p(x_{i}|x_1\ldots x_{i-1})=p(x_i|x_{i-1})$

对 $\forall i\in[1,n]$ 都成立，

故 $H(X_1\ldots X_n)=H(X_1)+H(X_2|X_1)+\ldots+H(X_n|X_{n-1})$

同理：$H(X_2 \dots X_n) = H(X_2) + H(X_3 | X_2) + \dots + H(X_n | X_{n-1})$

故 $I(X_1; X_2 \dots X_n) = H(X_1) + H(X_2 \dots X_n) - H(X_1 \dots X_n)$

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

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

$= I(X_2; X_1)$.