4.4相对熵

熵

熵在物理中表示的是一个系统的混乱程度

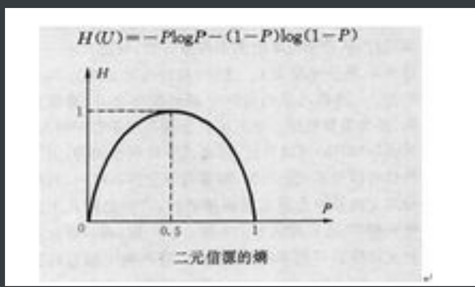
信息熵

表示的是某个概率分布的不确定程度

e g. 当一个概率分布的概率为99%则她的不确定性就越小，相应的信息熵越小

当这个事件发生的概率为50%时她的不确定性就大，则相应的信息熵就更大

信息熵的值域（0，1）



公式

The entropy $H(P)$ of a distribution $P(X)$

$$H(P) = -\sum_{x \in X} P(x) \log_b P(x)$$

其中x表示随机变量可能的取值，与p(x)表示的是x发生的概率 b是底数一般用2和e eg.

X	0	1
P(x)	0.4	0.6

$$
\begin{aligned}
H(P) &= -P(X=0)\log_2 P(X=0) - P(X=1)\log_2 P(X=1) \\
&= -0.4 * log_2(0.4) - 0.6 * log_2(0.6) \\
&\approx 0.97
\end{aligned}
$$

相对熵

相对熵是啥

就是如果有两个单独的概率分布P(X)和P(X)，我们用相对熵来衡量这两个分布的差异

公式

$$
\begin{aligned}
D_{\mathrm{KL}}(P\|Q) &= \sum_{x\in X} P(x)\log\left(\frac{P(x)}{Q(x)}\right) \\
&= \sum_{x\in X} P(x)\log P(X) - \sum_{x\in X} P(x)\log Q(X) \\
&= -H(P) + H(P,Q)
\end{aligned}
$$

在机器学习分类任务中，P往往用来表示样本的真实分布，比如[1,0,0]表示当前样本属于第一类。Q用来表示模型所预测的分布，比如[0.7,0.2,0.1] 。

性质

（1）kl(P||Q) >= 0，无最大值 (2)不对称 KL(P||Q) != KL(Q||P) (3)不满足三角不等式
（4）DKL的值越小，表示q分布和p分布越接近

(2)的计算过程

X	0	1	2
P(x)	0.36	0.48	0.16
Q(x)	0.333	0.333	0.333

$$
\begin{aligned}
D_{\mathrm{KL}}(P\|Q) &= \sum_{x\in X} P(x)\ln\left(\frac{P(x)}{Q(x)}\right) \\
&= 0.36\ln\left(\frac{0.36}{0.333}\right) + 0.48\ln\left(\frac{0.48}{0.333}\right) + 0.16\ln\left(\frac{0.16}{0.333}\right) \\
&= 0.0852996
\end{aligned}
$$

$$
\begin{aligned}
D_{\mathrm{KL}}(Q\|P) &= \sum_{x\in X} Q(x)\ln\left(\frac{Q(x)}{P(x)}\right) \\
&= 0.33\ln\left(\frac{0.333}{0.36}\right) + 0.333\ln\left(\frac{0.333}{0.48}\right) + 0.33\ln\left(\frac{0.333}{0.16}\right) \\
&= 0.097455
\end{aligned}
$$

python中kl散度咋弄

```python
import numpy as np
import scipy.stats

p=np.asarray([0.65,0.25,0.07,0.03])
q=np.array([0.6,0.25,0.1,0.05])

#方法一：根据公式求解
kl1=np.sum(p*np.log(p/q))
print(kl1)

#方法二：调用scipy包求解
kl2=scipy.stats.entropy(p, q)
print(kl2)
```