20131910023-金洋-IT-EX01

查阅资料， 思考并回答下述问题：

1 . 以公理化方式证明，一个随机事件发生的概率若为,其自信息具有的形式,其中C 为常数.

证：该自信息形式满足以下假设条件(其中C >0为常数)：

(1)单调性假设（可能性小的事件，拥有更大的信息量）：当时，易知，即是的递减函数;

(2)可加性假设(独立事件的信息量可以相加): 对于独立事件,, 有,则 (3)边界性假设: 当=0时, ; 当=1时, ;

2 . 用两种不同的观点解释熵.( 提示:事前推测结果的不确定性, 事后计算事件发生带来的信息)

答：事件未发生——熵表示事件发生的不确定性；

事件发生了——不确定性被消除了，熵表示事件提供的信息量；

3 . 熵,联合熵,条件熵,互信息之间存在什么关系.

答：(1)熵和联合熵：熵是随机变量的不确定度的度量，设离散型随机变量X~p(x),则X的熵定义为.

联合熵是将上述定义推广到多个随机变量的情形，以两个为例。对于服从联合分布为p(x,y)的一对离散型随机变量(X,Y),其联合熵定义为

.

1. 条件熵：是一个随机变量在给定另一个随机变量下的条件熵，若(X,Y)~p(x,y),则条件熵

.

1. 联合熵、条件熵之间的链式法则：.
2. 熵和互信息：，对称地也有；

即随机变量与自身的互信息为该随机变量的熵。故有时熵被称为自信息。

1. 熵、互信息、条件熵：，将代入，故有

.

4 . H(Y|X)=0 的条件是什么?

答：当Y是完全确定事件时；

理解一：当Y是完全确定事件时，Y的不确定性已经消除，X的发生与否对Y的不确定性没有影响，故H(Y|X)=0；

理解二：由链式法则H(Y|X)=H(X,Y)-H(X);当且仅当H(X,Y)=H(X)，H(Y|X)=0；

而H(X,Y)=H(X)即表示事件Y为确定事件.

5 . I(Y;X)=0 的条件是什么?

答：当X,Y为独立事件时；

∵ I(Y;X)=H(Y)-H(Y;X);

仅当H(Y)=H(Y;X)时，上式=0，即需Y，X为独立事件.

6 . I(X;X)=0 的条件是什么?

答：当H(X)=0时，即事件X为完全确定事件.

7 . H(X)=0 的条件是什么?

答：X事件是完全确定事件，例如对于0-1分布事件，p(X=0)=0,p(X=1)=1时该事件完全确定，没有不确定性，故熵为0；

8 . 将相对熵作为距离对待, 与距离公理相对照,存在哪些差异?

答：距离公理需要满足①非负 ②对称 ③△不等式；

而相对熵总是非负的（当且仅当p=q时为0），但相对熵并不对称，也不满足三角不等式。

对不满足三角不等式验证如下：若有三个分布A,B,C且A(0)=1/2, A(1)=1/2 ; B(0)=1/4,B(1)=3/4 ;C(0)=1/8,C(1)=7/8 , 则

D(A||C) - (D(A||B)+ D(B||C)) =1/2log2+1/4log(7/6)>0,

即相对熵不满足三角不等式。

9 . 从关系数据库的角度看, 如果一个表(table) 中存在多个属性列, 每列看成变量及其取值,研究任意两列(C1,C2)之间的关系:

如果H(C1|C2)=0,表明什么?

如果I(C1;C2)=0,表明什么?

答：如果H(C1|C2)=0，表明在C2确定时，C1也被确定下来，即C2是候选码

如果I(C1;C2)=0,表明在给出C2（或C1）的情况下，对另一个属性变量的取值不确定性缩减量为0，即C1和C2都不是候选码。

1. 从二值分布的熵函数H(p) 的取值进行观察,如果在多值分布的情况下考虑熵函数H(p1 ,..., pi ,...,pn )它的最大值在什么情形下可以达到?进一步,证明你的结论.

答：H(p1 ,..., pi ,...,pn )≤logn，当且仅当X服从均匀分布时，即时，等号成立.

证法一：设随机变量X，是样本空间，是上均匀分布的概率密度函数，p(x)是X的密度函数，则,

又由于相对熵的非负性，则.

∴，当X为均匀分布时，可以取等号。

证法二：



又对于x>0,则有（当且仅当x=1时等号成立），两侧同乘log(e)后，有

故



当且仅当，——当X为均匀分布时，可以取等号。

（ 答案提交方式： 电子稿， 电子邮箱）