鉴别信息

2020年4月2日 11:29

1. 定义

对于一个随机变量可能会存在不同分布的可能性，信息熵之间的差别称需要一种度量。这个度量称之鉴别信息，定义如下。

两个随机分布 $p(x)$ 和 $q(x)$ 之间鉴别信息的定义为

$$D(p\|q)=\sum_{x\in X}p(x)\log\frac{p(x)}{q(x)}$$

这与对数求和不等式的形式是一致的。这是一个下凸函数。当 $p(x)=q(x)$ 时，鉴别信息的值为0，则两者的距离相近。

考虑一下距离度量的函数要求

① 作为距离函数应是非负的

② 从两者出发距离要相同

③ 满足三角不等式，两边之和大于第三边。

而鉴别信息只满足第一点。因此不能严格的称其为距离。那么虽然这也让鉴别信息的应用受到限制。但是他与熵和互信息存在着本质的联系。因为信息熵也不是对称。

2. 熵、互信息、鉴别信息的关系

先讨论熵和鉴别信息的关系。他们满足

$$H(X)=\log|X|-D(p\|u)$$

其中 $\log|X|$ 是最大熵，在分布为u时取得。那么我们知道这是在等概分布上取得的，那么 $D(p\|u)$ 是均匀分布与实际分布之间的差异的度量

证明：

$$H(x)=\log N-\log N-\sum_{x\in X}p(x)\log p(x)$$

$$=\log|X|+\sum_{x\in X}p(x)\log\frac{1}{N}-\sum p(x)\log p(x)$$

$$=\log|X|-\sum_{x\in X}p(x)\log\frac{u}{p(x)}$$

于是得证。

从这里我们也可以说明鉴别信息是一个下凸函数。

再讨论鉴别信息和两个随机变量的互信息间的关系

$$I(x;y)=D(p(x,y)\,\|\,p(x)p(y)).$$

从定义就可以知道

$$I(x,y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

于是得证。

从这点上说，信息熵就是等概分布的信息熵减去他们之间的距离，互信息则是独立减去联合，这也说明了他们之间的内在联系。

3. 鉴别信息的凸性

关于他的下凸特性，可以直接参考对数求和不等式的下凸证明过程