互信息

Introduction

- **信息量**是对某个事件发生或者变量出现的概率的度量

  一般一个事件发生的概率越低，事件包含的信息量越大，这跟我们直观上的认知也是吻合的，越稀奇新闻包含的信息量越大，因为这种新闻出现的概率低，香农提出了一个定量衡量信息量的公式 $\log \frac{1}{p} = -\log p$

- **熵**是衡量一个系统的稳定程度，其实就是一个系统所有变量信息量的期望或者说是均值

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

- **互信息**是信息论中评价两个随机变量之间依赖程度的一个度量

  举个例子：x=今天下雨与y=今天阴天，显然在已知y的情况下，发生x的概率会更大

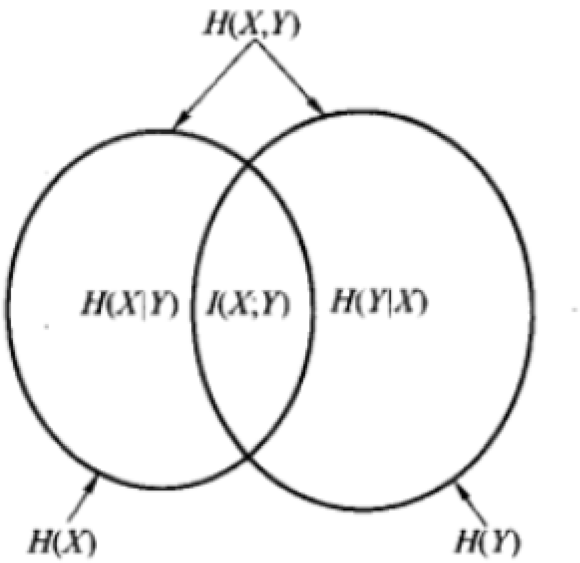
  对于两个随机变量X和Y，其联合分布为$p(x,y)$, 边缘分布为$p(x), p(y)$，则互信息可以定义为：

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

  推导过程：

$$
\begin{aligned}
I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(y)p(x|y)}{p(x)p(y)} \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x|y)}{p(x)} \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x|y) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x) \\
&= -\sum_{x\in\mathcal{X}} p(x)\log p(x) - [-\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x|y)] \\
&= H(X) - H(X|Y)
\end{aligned}
$$

根据推到结果可知，**互信息I（$X;Y$）表示知道事实Y后，原来信息量减少了多少**



reference

[1]blog https://zhuanlan.zhihu.com/p/36192699