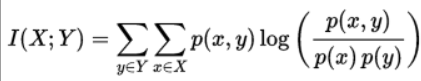
李沐老师的《动手学深度学习》

# 互信息：

度量两个集合间的相关性（点间互信息的期望值，单位bit）

## 公式：

离散随机变量x，y的互信息定义（其中p(x,y) 是X和Y的[联合概率分布函数](https://zh.wikipedia.org/wiki/%E8%81%94%E5%90%88%E5%88%86%E5%B8%83)，而p(x)和p(y)分别是X和Y的[边缘概率](https://zh.wikipedia.org/wiki/%E8%BE%B9%E7%BC%98%E6%A6%82%E7%8E%87)分布函数。）



在[连续随机变量](https://zh.wikipedia.org/wiki/%E8%BF%9E%E7%BB%AD%E5%87%BD%E6%95%B0)的情形下（求和被替换成了[二重定积分](https://zh.wikipedia.org/wiki/%E4%BA%8C%E9%87%8D%E7%A7%AF%E5%88%86)，其中p(x,y) 当前是X和Y的联合概率密度函数，而p(x)和p(y)分别是X和Y的边缘概率密度函数。）



## 互信息量I(xi;yj)在联合概率空间P(XY)中的统计平均值

平均互信息I(X;Y)克服了互信息量I(xi;yj)的随机性,成为一个确定的量。如果对数以2为基底，互信息的单位是[bit](https://zh.wikipedia.org/wiki/%E4%BD%8D%E5%85%83)

直观上，互信息度量X和Y共享的信息：它度量知道这两个变量其中一个，对另一个不确定度减少的程度。例如，如果X和Y相互独立，则知道X不对Y提供任何信息，反之亦然，所以它们的互信息为零。在另一个极端，如果X是Y的一个确定性函数，且Y也是X的一个确定性函数，那么传递的所有信息被X和Y共享：知道X决定Y 的值，反之亦然。因此，在此情形互信息与 Y（0或 X）单独包含的不确定度相同，称作 Y（或 X）的[熵](https://zh.wikipedia.org/wiki/%E4%BF%A1%E6%81%AF%E7%86%B5)。（？）而且，这个互信息与 X 的熵和 Y 的熵相同。（这种情形的一个非常特殊的情况是当 X 和 Y 为相同随机变量时。）

### 互信息的性质

1. 是非负的（即 I(X;Y) ≥ 0）
2. 是[对称的](https://zh.wikipedia.org/w/index.php?title=%E5%AF%B9%E7%A7%B0%E5%87%BD%E6%95%B0&action=edit&redlink=1)【由Y提取到的关于X的信息量与从X中提取到的关于Y的信息量是一样的。】（即 I(X;Y) = I(Y;X)）。
3. I(X; Y) = 0 [当且仅当](https://zh.wikipedia.org/wiki/%E5%BD%93%E4%B8%94%E4%BB%85%E5%BD%93) X 和 Y 为独立随机变量

（推导）当 X 和 Y 独立时，p(x,y) = p(x) p(y)，因此：

1. I(X;Y)≤H(X)，I(Y;X)≤H(Y)

（极值性）

从一个事件提取关于另一个事件的信息量, 至多是另一个事件的熵那么多, 不会超过另一个事件自身所含的信息量。

当X和Y是一一对应关系时: I(X;Y)=H(X), 这时H(X/Y)=0。

当X和Y相互独立时: H(X/Y) =H(X), I(Y;X)=0。 从一个事件不能得到另一个事件的任何信息,这等效于信道中断的情况。

### 5. 凸函数性

　　平均互信息量是p(xi)和p(yj /xi)的函数,即I(X;Y)=f [p(xi), p(yj /xi)];

　　若固定信道,调整信源, 则平均互信息量I(X;Y)是p(xi)的函数,即I(X;Y)=f [p(xi)];

　　若固定信源,调整信道, 则平均互信息量I(X;Y)是p(yj /xi)的函数,即I(X;Y)=f [p (yj /xi)]。

　　平均互信息量I(X;Y)是输入信源概率分布p(xi)的上凸函数(concave function; or convext cap function)。

　　平均互信息量I(X;Y)是输入转移概率分布p(yj /xi)的下凸函数(convext function; or convext cup function)。

李璐学姐

自然语言处理nlp

多任务处理

提示初始化

提示符号与诱导图关系（符号维度与节点特征维度）

Aij可调维度

（提示符号点集）

预训练模型

节点级，编辑，任务操作——>图集任务

重叠子空间

诱导图（代替节点）重构

图分裂任务

点分裂任务

邻居

N跳子图

编带权最短路径

* Translation平移不变性：Translation Invariance

Kg

Tell

头向量，尾向量

元学习（通过模型测试初始化参数【知识级（训练【编码】），查询集（检验【解码】）】）多任务集

学会学习

直觉 learning to learn

损失函数优化（神经网路）

Maining test

弱人工智能 强人工智能 aji

Agrugaition

冻结

大语言模型

下游任务

Inprudution

Ld（f）

梯度下降

Gn

时间复杂度

（多方面证明）

第三个方面

兼容性（不需要任务头，不需要下游处理）

Graph cl

可解释性

1.

2．适应性

3.影响

4.和其他方法作比较

5.掌控图的能力

监督学习

消融参数

2.佳乐

邻居，i j的邻居

Pdr

Mlp多层感知器

过渡平滑

Soft max 损失函数

[交叉熵损失](https://so.csdn.net/so/search?q=%E4%BA%A4%E5%8F%89%E7%86%B5%E6%8D%9F%E5%A4%B1&spm=1001.2101.3001.7020)来衡量模型预测结果和真实标签之间的差距

交叉熵从P到Q，记为H(P,Q)。我们可以把交叉熵想象为“主观概率为Q的观察者在看到根据概率P生成的数据时的预期惊异”。  
当P=Q时，交叉熵达到最低。 在这种情况下，从P到Q的交叉熵是H(P,P)=H(P )。

Digrae