# CrossEntropy

香农认为：信息是用来消除不确定性的东西，不确定性的数学表达可以转化为概率的形式

信息越少，可能性越大

## 信息熵

把信息中排出了冗余后的平均信息量称为“信息熵”，设表示样本空间的第i个样本，其概率为，样本空间总共有n个样本，信息熵的公式

相关性质：

极值性：在各个样本等概率出现的条件下取最大值，也就是说明等概率的时候最为混乱，而且样本数据越多，熵越大

连续性

对称性（交换样本顺序，结果不变）

## 交叉熵

交叉熵=KL散度+信息熵

KL散度的公式：

交叉熵的公式：

对于二分类和多分类的交叉熵公式如下：这是因为在二分类中对用的网络最后一层是sigmoid，多分类中是softmax：

我们知道KL散度可以衡量两个分布之间的差异性，对于一个样本空间而言，其信息熵是确定的，因此也可以用交叉熵衡量两个分布的差异

