交叉熵 相对熵（KL散度/互熵）

Arya鑫 (/u/1d11532897bb) + 关注
2017.08.03 11:36* 字数 723 阅读 488 评论 0 喜欢 4
(/u/1d11532897bb)

香农熵

熵考察（香农熵）的是单个的信息（分布）的期望：反映了一个系统的无序化（有序化）程度，一个系统越有序，信息熵就越低，反之就越高。

$$H(p) = -\sum_{i=1}^{n} p(x_i) \log p(x_i)$$

交叉熵

交叉熵考察的是两个的信息（分布）的期望：

交叉熵和熵，相当于，协方差和方差

$$H(p, q) = -\sum_{i=1}^{n} p(x_i) \log q(x_i)$$

相对熵

相对熵考察两个信息（分布）之间的**不相似性：**

所谓相对，自然在两个随机变量之间。又称互熵，Kullback–Leibler divergence（K-L散度）等。设p(x)和q(x)是X取值的两个概率分布，则p对q的相对熵为：

$$KL(p\|q) = -\sum_x p(x) \log \frac{q(x)}{p(x)}$$

在一定程度上，**熵可以度量两个随机变量的距离。**KL 散度是两个概率分布 P 和 Q 差别的**非对称性的度量。**KL 散度是用来度量使用基于 Q 的编码来编码来自 P 的样本平均所需的额外的位元数。

典型情况下，**P 表示数据的真实分布**，Q 表示数据的**理论分布**，模型分布，或 P 的近似分布。

相对熵的性质，相对熵（KL散度）有两个主要的性质。如下

（1）尽管 KL 散度从直观上是个度量或距离函数，但它并不是一个真正的度量或者距离，因为它**不具有对称性**，即

$$D(p\|q) \neq D(q\|p)$$

（2）相对熵的值为非负值，即

$$D(p\|q) \geq 0$$

三者之间的关系：

$$\begin{aligned} H(p, q) &= -\sum_x p(x) \log q(x) \\ &= -\sum_x p(x) \log p(x) - \sum_x p(x) \log \frac{q(x)}{p(x)} \\ &= H(p) + KL(p\|q) \end{aligned}$$

简森不等式与 KL散度：

$$KL(p\|q) = -\int p(x) \ln \frac{q(x)}{p(x)}\, dx$$

因为−lnx是凸函数，所以满足，凸函数的简森不等式的性质：

$$f(\mathbb{E}) \leq \mathbb{E}(f)$$

这里我们令f(·)=−lnx，则其是关于x的凸函数，因此：

也即 KL 散度恒大于等于 0；

联合熵

联合熵:（X,Y）在一起时的不确定性度量

$$H(X,Y) = -\sum_{x,y} p(x,y) lnp(x,y)$$

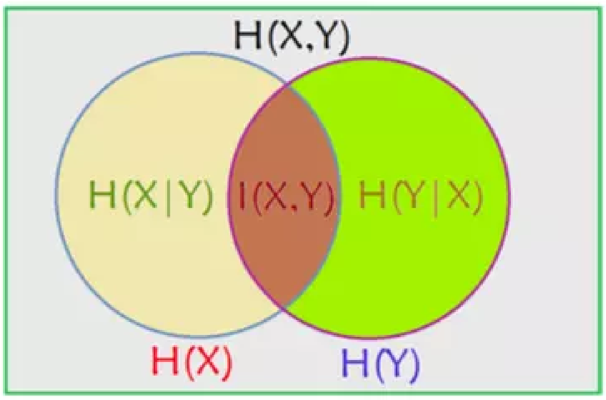
条件熵

条件熵:　X确定时，Y的不确定性度量

在X发生是前提下，Y发生新带来的熵。

$$H(X,Y) - H(X) = -\sum_{x,y} p(x,y) lnp(y|x)$$

联系：



?

如果是回归问题的，使用平方损失函数。分类问题建议使用交叉熵损失，用平方损失会出现如下问题：在误差较大时，损失函数比较平坦，更新较慢，还会出现梯度为0的情况，期望的情况是训练完成时，到达某个极值点，这时梯度为0，所以就没办法判断训练是否完成了。交叉熵损失就不会出现这种情况，在远离目标的时候，曲线比较陡峭。

来源：http://blog.csdn.net/lanchunhui/article/details/50970625

http://blog.csdn.net/lanchunhui/article/details/53365438

http://blog.csdn.net/lanchunhui/article/details/51277608

http://www.cnblogs.com/little-YTMM/p/5582271.html