sklearn：点互信息和互信息 - 专注计算机体系结构 - CSDN博客

原 sklearn：点互信息和互信息

2017年06月03日 00:07:53 JepsonWong 阅读数 10131

1、点互信息PMI

机器学习相关文献里面，经常会用到点互信息PMI(Pointwise Mutual Information)这个指标来衡量两个事物之

词）。

其原理很简单，公式如下：

$$PMI(x;y) = log\frac{p(x,y)}{p(x)p(y)} = log\frac{p(x|y)}{p(x)} = log\frac{p(y|x)}{p(y)}$$

72848755
159.1 KB

在概率论中，我们知道，如果x跟y不相关，则p(x,y)=p(x)p(y)。二者相关性越大，则p(x, y)就相比于p(x)p(y)越

好理解，在y出现的情况下x出现的条件概率p(x|y)除以x本身出现的概率p(x)，自然就表示x跟y的相关程度。

举个自然语言处理中的例子来说，我们想衡量like这个词的极性（正向情感还是负向情感）。我们可以预先挑选

比如good。然后我们算like跟good的PMI。

2、互信息MI

点互信息PMI其实就是从信息论里面的互信息这个概念里面衍生出来的。

互信息即：

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

72848755
151.5 KB

其衡量的是两个随机变量之间的相关性，即一个随机变量中包含的关于另一个随机变量的信息量。所谓的随机变

果的量的表示，可以简单理解为按照一个概率分布进行取值的变量，比如随机抽查的一个人的身高就是一个随机

可以看出，互信息其实就是对X和Y的所有可能的取值情况的点互信息PMI的加权和。因此，点互信息这个名字

3、sklearn编程

```
from sklearn import metrics as mr
mr.mutual_info_score(label,x)
```