### 最大熵模型中的对数似然函数表示法解释



5 人赞同了该文章

书中最大熵模型的极大似然估计这部分中,条件概率分布的对数似然函数表示为

$$L_{\tilde{P}}(P_w) = log \prod_{x,y} P(y|x)^{\tilde{P}(x,y)} = \sum_{x,y} \tilde{P}(x,y) log P(y|x)$$
 (1)

感觉似然函数应该表示为:

$$L_{ ilde{P}}(P_w) = log \prod_{x,y} P(y|x)$$
 (2)

但是式 (1) 中似然函数出现了指数形式。

其实公式 (2) 是错误的。似然函数定义为样本集中各个样本的联合概率,最大熵模型的训练数据集  $T=\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$ ,因此,似然函数应该为:

$$L_{ ilde{P}}(P_w) = log \prod_{i=1}^n P(x_i, y_i)$$
 (3)

博文《最大熵模型中的对数似然函数的解释》给出了解释,推导过程如下:

训练数据集 T 中包含n个样本,  $(x_i,y_i)$  为其中一个样本。样本集中会有很多值相同的样本,即  $(x_i,y_i)=(x_j,y_j)(i\neq j)$  。假设样本的取值为k个,分别为  $\{(v_1,w_1),(v_2,w_2),\ldots,(v_k,w_k)\}$ ,用  $C[(X,Y)=(v_i,w_i)]$  表示训练集中样本值为  $(v_i,w_i)$  的频数,则公式(3)可以表示为:

$$L_{ ilde{P}}(P_w) = log \prod_{i=1}^k P(v_i, w_i)^{C[(X,Y) = (v_i, w_i)]}$$
 (4)

等号两边同时开n次方,可得

$$L_{ ilde{P}}(P_w)^{rac{1}{n}} = log \prod_{i=1}^k P(v_i, w_i)^{rac{C[(X,Y) = (v_i, w_i)]}{n}}$$
 (5)

而经验概率分布  $ilde{P}(X=v_i,Y=w_i)=rac{C[(X,Y)=(v_i,w_i)]}{n}$  , 所以公式 (5) 可表示为

$$L_{ ilde{P}}(P_w)^{rac{1}{n}} = log \prod_{i=1}^k P(v_i,w_i)^{ ilde{P}(v_i,w_i)}$$
 (6) 博文《最

大熵模型中的对数似然函数的解释》中描述说将公式 (6) 简化写成

$$L_{\tilde{P}}(P_w)^{\frac{1}{n}} = \log \prod_x P(x, y)^{\tilde{P}(x, y)} \tag{7}$$

这个地方有待商榷,因为公式 (7) 中连乘是针对n个样本,6式中连乘是针对k个样本值,一般 k < n ,显然两式的结果不一样,公式7的结果会小于6的结果。

浙江工业大学教师Jerry在回答知乎提问《为什么最大熵模型的极大似然估计中带有指数》时,提出引用交叉熵的概念解释进行解释,解释如下:

**交叉熵(CrossEntropy)**用来度量真实分布 P(x,y) 与经验分布  $\tilde{P}(x,y)$  在(X,Y)的差异性,我们希望经验分布于真实分布之间的差异最小化,即**交叉熵最小**。

$$egin{aligned} CrossEntropy &= -\log \prod_{x,y} p(x,y)^{\hat{p}(x,y)} \ &= -\sum_{x,y} \hat{p}(x,y) \log p(x,y) \ &= -\sum_{x,y} \hat{p}(x,y) \log[\hat{p}(x)p(y|x)] \ &= -\sum_{x,y} \hat{p}(x,y) \log p(y|x) + \sum_{x,y} \hat{p}(x,y) \log \hat{p}(x) \end{aligned}$$

由于公式(8)结果中的第2项为常数,所以极小化CrossEntropy等价于极大化  $\sum_{x,y} \tilde{P}(x,y)logP(y|x)$  ,即公式(1)中对数似然函数,这个就是书中为什么出现指数型似然函数的原因。

显然这个解释更有说服力。

## 参考文献

【1】最大熵模型中的对数似然函数的解释

【2】为什么最大熵模型的极大似然估计中带有指数

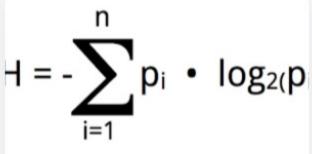
发布于 2019-09-24



## 文章被以下专栏收录

统计学习方法笔记

## 推荐阅读



信息熵数值计算

Dreisteine



最大熵模型与最大似然估计

南京汤包没有馅

# 信息论 (自信息 信息熵 联合熵条件熵 互信息)

1、约定数学表达形式大写 X \quad Y \quad Z 表示随机变量小写 x \quad y \quad z 表达随机变量具体 取值花体 \mathcal{X} \quad \mathcal{Y} \quad \mathcal{Z} 表 达集合集合的势,即集合中…

xxids

傲慢仍

## 最小二乘法和最大似然法异同

请参考以下链接:
http://blog.sina.com.cn/s/blog\_4b
http://blog.csdn.net/luo86106/arti
对于最小二乘法,当从模型总体随机
抽取n组样本观测值后...

傲慢偏见

