数学 统计学 机器学习 最大熵

关注者 32 被浏览 3,152

为什么最大熵模型的极大似然估计中带有指数?

关于最大熵模型的极大似然估计有一点不太理解，为什么待求解模型的似然函数中带有指数？具体是怎么推导出来的？

关于条件分布 $P(Y|X)$ 的熵为：

$$H(P) = -\sum_{x,y} P(y,x) log P(y|x) = -\sum_{x,y} \tilde{P}(x) P(y|x) log P(y|x)$$

$\tilde{P}(x)$ 和 $\tilde{P}(x,y)$ 是经验分布：

$$\tilde{P}(X=x,Y=y) = \frac{count(X=x,Y=y)}{N}$$

$$\tilde{P}(X=x) = \frac{count(X=x)}{N}$$

待求解的概率模型 $P(Y|X)$ 的似然函数为：

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

(最大熵模型 Maximum Entropy Model)

关注问题 写回答 邀请回答 好问题 1 3 条评论 分享 ··· 收起

3 个回答

默认排序

zero
计算机硕士

14 人赞同了该回答

最大熵模型中的对数似然函数的解释

发布于 2017-09-13

▲ 赞同 14 ▼ 1 条评论 分享 收藏 喜欢

知乎用户

8 人赞同了该回答

下面阐述不讲究严谨的数学，只关注idea的解释。

首先理解下**二分类**的逻辑回归的交叉熵似然 或 loss (两者差一个负号)， 这里的 y_i 即为 x_i 样本的输出，而 $p(x_i)$ 为模型的输出。

$$L(w) = \prod [p(x_i)]^{y_i} [1-p(x_i)]^{1-y_i}$$

写成**多类别**的交叉熵即为

$$L(w) = \prod [p(x_i)]^{y_i}$$

注： $y_i = [0\ 0\ 0\ 1\ \dots\ 0]$ 为one-hot向量，表示类别。交叉熵的本质可以用来度量两个分布的差异性。

最大熵模型的似然是使用了(模型学的)真实分布 $p(x,y)$ 与(来自数据的)经验分布 $\hat{p}(x,y)$ 的**交叉熵来定义**。

注：知识迁移，逻辑回归的交叉熵似然---》最大熵模型的交叉熵似然。

对于样本(X,Y)，它的似然使用交叉熵定义：[用来度量真实分布与经验分布在(X,Y)的差异性 (目标当然希望真实分布与经验分布越接近也好了)]

$$\begin{aligned} L_{\hat{p}} &= \prod_{x,y} p(x,y)^{\hat{p}(x,y)} \\ &= \prod_{x,y} [\hat{p}(x)p(y|x)]^{\hat{p}(x,y)} \\ &= \prod_{x,y} [p(y|x)]^{\hat{p}(x,y)} \times \prod_{x,y} [\hat{p}(x)]^{\hat{p}(x,y)} \end{aligned}$$

注：模型使用 $\hat{p}(x) \approx p(x)$,因此有 $p(x,y) = p(x)p(y|x) = \hat{p}(x)p(y|x)$

取log似然有(上面式子再重新推一遍)

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

注：凡是带 ^ 都表示已知的，来自数据的。所以，第二项是常量，在对最大似然优化无贡献，直接舍去。

$$L_{\hat{p}} = \sum_{x,y} \hat{p}(x,y) \log p(y|x)$$

即为最大熵模型的似然函数。

转载请注明出处。

编辑于 2019-09-25

▲ 赞同 8 ▼ 2 条评论 分享 收藏 喜欢 收起

长行

因为在训练数据集中，(x,y) 的样本可能不止一个，而连乘符号中仅区分了不同的 (x,y)。如果没有指数的话，那么无论 (x,y) 的出现频数是多少，均只乘了一次。

$\tilde{P}(x,y)$ 正是 (x,y) 的样本的频数，因此使用 $\tilde{P}(x,y)$ 作为指数，以表示 (x,y) 的样本需要连乘的次数。

发布于 06-15

▲ 赞同 ▼ 添加评论 分享 收藏 喜欢

写回答

1 个回答被折叠（为什么？）