最大熵模型与Logistic回归

www.mozheyang.top

最大熵模型

“知之为知之，不知为不知，是知也”

在保留模型最大不确定性的前提下

将模型不确定性降到最低

最大熵模型

Maximum entropy model

一个看起来有点矛盾的假设

定义

特征函数：描述输入x与输出y之间的某个事实

$$f(x,y)=\begin{cases}1, & if\ criteria\ of\ x\ and\ y\ matched\\ 0, & else\ not\ matched\end{cases}$$

一般地，特征函数可以是任意实值函数

定义

联合概率的经验分布 $$\tilde{P}(x,y)=\tilde{P}(X=x,Y=y)=\frac{v(X=x,Y=y)}{N}$$

边缘概率的经验分布 $$\tilde{P}(x)=\tilde{P}(X=x)=\frac{v(X=x)}{N}$$

概率与频率的关系

频率容易从实际数据中获得

经验分布是基于实际观测数据构建的分布，用于估计总体分布

定义

贝叶斯定理

Bayes' Theorem

$$P(x,y) = {\color{red}P(y|x)}P(x) = P(x|y)P(y)$$

条件概率分布 ${\color{red}P(y|x)}$ 建模了观测x向标签y的映射规律，为我们所需要求解的模型

如果模型反映了现实世界?

那么真实世界的xy分布应该与模型预测的xy分布相同

➡ 两个分布的n阶矩相等

Moment Theorem

特殊情况下为充要条件

➡ 两个分布的一阶矩（期望）相等

➡ $$\sum_{x,y} \tilde{P}(x,y) f(x,y) = \sum_{x,y} \tilde{P}(x) P(y|x) f(x,y)$$

真实世界分布的期望$E_P(f)$　　模型预测分布的期望$E_{\tilde{P}}(f)$

最大熵模型要最大化什么“熵”？

最大化模型集合的条件“熵”

给定x的前提下，模型集合要保留最大的不确定性

模型集合的条件熵

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

约束最优化问题

最大化模型集合的条件熵

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

$$subject\ to \begin{cases} E_P(f_i) - E_{\tilde{P}}(f_i) = 0, i = 1,2, \dots n \\ \sum_y P(y|x) = 1 \end{cases}$$

约束最优化问题

最大化变成最小化问题

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

$$subject\ to \begin{cases} E_P(f_i) - E_{\tilde{P}}(f_i) = 0, i = 1,2,\ldots n \\ \sum_{y} P(y|x) = 1 \end{cases}$$

约束最优化问题

引入拉格朗日乘子ω_i，将有约束问题变为无约束问题

$$L(P,\boldsymbol{\omega}) = -H(P) + \omega_0\left(1 - \sum_{y} P(y|x)\right) + \sum_{i=1}^{n} \omega_i\left(E_P(f_i) - E_{\tilde{P}}(f_i)\right)$$

$$\begin{aligned} L(P,\boldsymbol{\omega}) &= \sum_{x,y} \tilde{P}(x)P(y|x)\log P(y|x) + \omega_0\left(1 - \sum_{y} P(y|x)\right) \\ &+ \sum_{i=1}^{n} \omega_i\left(\sum_{x,y} \tilde{P}(x,y)f_i(x,y) - \sum_{x,y} \tilde{P}(x)P(y|x)f_i\,(y|x)\right) \end{aligned}$$

约束最优化问题

对拉格朗日函数求极值，令偏导数为0

$$L(P,\boldsymbol{\omega})=\sum_{x,y}\tilde{P}(x)P(y|x)\log P(y|x)+\omega_0\left(1-\sum_{y}P(y|x)\right)+\sum_{i=1}^{n}\omega_i\left(\sum_{x,y}\tilde{P}(x,y)f_i(x,y)-\sum_{x,y}\tilde{P}(x)P(y|x)f_i(y|x)\right)$$

$$\frac{\partial L(P,\boldsymbol{\omega})}{\partial P(y|x)}=\sum_{x,y}\tilde{P}(x)(\log P(y|x)+1)-\sum_{y}\omega_0-\sum_{i=1}^{n}\omega_i\sum_{x,y}\tilde{P}(x)f_i(y|x)$$

$$=\sum_{x,y}\tilde{P}(x)(\log P(y|x)+1)-\sum_{x,y}\tilde{P}(x)\omega_0-\sum_{x,y}\tilde{P}(x)\sum_{i=1}^{n}\omega_i f_i(y|x)$$

$$=\sum_{x,y}\tilde{P}(x)\left(\log P(y|x)+1-\omega_0-\sum_{i=1}^{n}\omega_i f_i(y|x)\right)=0$$

约束最优化问题

若极值存在， 且解对$\tilde{P}(x)$的变化不敏感，则每项均为0

$$\log P(y|x) + 1 - \omega_0 - \sum_{i=1}^{n} \omega_i f_i(y|x) = 0$$

$$\Rightarrow \quad P(y|x) = \exp\left(\omega_0 - 1 + \sum_{i=1}^{n} \omega_i f_i(y|x)\right)$$

$$= \exp(\omega_0 - 1) \cdot \exp\left(\sum_{i=1}^{n} \omega_i f_i(y|x)\right)$$

$$= \frac{1}{Z_\omega(x)} \cdot \exp\left(\sum_{i=1}^{n} \omega_i f_i(y|x)\right)$$

约束最优化问题

模型集合概率满足积分为1的条件

$$\sum_{y} P(y|x) = \sum_{y} \frac{1}{Z_{\omega}(x)} \cdot \exp\left(\sum_{i=1}^{n} \omega_i f_i(y|x)\right) = 1$$

$$\Rightarrow \quad Z_{\omega}(x) = \sum_{y} \exp\left(\sum_{i=1}^{n} \omega_i f_i(y|x)\right)$$

最大熵模型的形式

模型集合概率满足积分为1的条件

$$P(y|x) = \frac{\exp(\sum_{i=1}^{n} \omega_i f_i(y|x))}{\sum_y \exp\left(\sum_{i=1}^{n} \omega_i f_i(y|x)\right)}$$

也即softmax函数的形式

$$f(x) = \frac{e^x}{\sum_i e^{x_i}} \quad \Longleftrightarrow \quad f(x) = \frac{1}{1+e^x} = \frac{e^0}{e^0+e^x}$$

与sigmoid函数形式也是相似的

与Logistic回归的关系?

Logistic回归又被称为对数概率回归

$$\log \frac{p}{1-p} = AX \quad \Rightarrow \quad p = \frac{1}{1+e^{AX}}$$

更严肃的表达

$$P(y=1|x) = \sigma(\omega^T x) = \frac{1}{1+e^{\omega^T x}}$$

这里其实就是sigmoid函数

与Logistic回归的关系?

对特征函数$f(x,y)=xy$来说

$$P(y|x)=\frac{\exp\left(\sum_{i=1}^{n}\omega_i f_i(y|x)\right)}{\sum_y \exp\left(\sum_{i=1}^{n}\omega_i f_i(y|x)\right)}=\frac{\exp\left(\sum_{i=1}^{n}\omega_i xy\right)}{\sum_y \exp\left(\sum_{i=1}^{n}\omega_i xy\right)}$$

$$P(y=0|x)=\frac{1}{1+\exp\left(\sum_{i=1}^{n}\omega_i x\right)}$$

$$P(y=1|x)=\frac{\exp\left(\sum_{i=1}^{n}\omega_i x\right)}{1+\exp\left(\sum_{i=1}^{n}\omega_i x\right)}$$

$$= \quad P(y=1|x)=\sigma(\omega^T x)=\frac{1}{1+e^{\omega^T x}}$$

可以认为Logistic回归是最大熵模型下的一个特例

THANKS