西南財經大學

SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS

1925 - 2021

经世济民 孜孜以求

机器学习基础

逻辑斯蒂回归

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

- 模型
- 总结

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

- 模型
- 总结

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

逻辑斯蒂分布

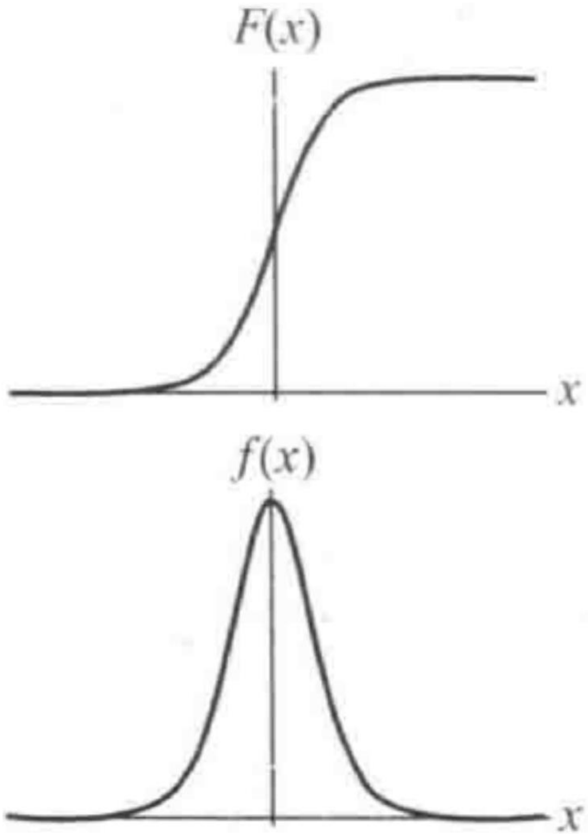
- 设 X 是连续随机变量
  - 分布函数

$$F(X) = P(X \leq x) = \frac{1}{1 + e^{-(x-\mu)/\gamma}}$$

  - 密度函数

$$f(x) = F'(x) = \frac{e^{-(x-\mu)/\gamma}}{\gamma(1 + e^{-(x-\mu)/\gamma})^2}$$

- 分布函数的图像是一条S形曲线 (sigmoid curve)



黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

逻辑斯蒂回归模型

- 二项逻辑斯蒂回归模型是一种分类模型
- 条件概率

$$P(Y=1|x)=1-F(w\cdot x)=\frac{\exp(w\cdot x)}{1+\exp(w\cdot x)}$$

$$P(Y=0|x)=F(w\cdot x)=\frac{1}{1+\exp(w\cdot x)}$$

- 对于给定输入样本x，计算以上两个条件概率，将样本分到概率值较大的那一类

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

几率（odds）

- 一个事件的几率是指该事件发生的概率与该事件不发生的概率的比值
- 逻辑斯蒂回归的对数几率 $\log\frac{P(Y=1|x)}{1-P(Y=1|x)}=w\cdot x$
  - 输出 $Y=1$ 的对数几率是输入 x 的线性函数。即用线性回归模型的预测结果去逼近真实标记的对数几率，因此，也叫对数几率回归模型（logistic regression）。

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

感知机与逻辑斯蒂回归

- 若 $w\cdot x\geq 0$ 分为正类等价于 $P(Y=1|x)\geq P(Y=0|x)$
  - 和感知机模型有同样的分类结果
- 都是线性分类器，前者作用于阶跃函数，后者是逻辑斯蒂函数
- 逻辑斯蒂回归有概率解释能力，函数光滑，有更好的分类结果
- 前者损失函数为距离损失函数，并且在有限步内收敛
- 后者损失函数由最大似然函数得到，需要设置最大迭代次数，否则梯度下降无法停止

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

参数估计

- 设 $P(Y=1|x)=\pi(x)$，$P(Y=0|x)=1-\pi(x)$，则似然函数为

$$\prod_{i=1}^{N}[\pi(x_i)]^{y_i}[1-\pi(x_i)]^{1-y_i}$$

- 对数似然函数为 $L(w)=\sum_{i=1}^{N}[y_i(w\cdot x_i)-\log(1+\exp(w\cdot x_i))]$
- 对 $L(w)$ 求极大值，得到 w 的估计（梯度下降、拟牛顿法）
- $L(w)$ 是上凸函数

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

多项逻辑斯蒂回归

- 设随机变量 Y 的取值集合为 $\{1,2,\cdots,K\}$，则模型为

$$P(Y=k|x)=\frac{\exp(w_k\cdot x)}{1+\sum_{k=1}^{K-1}\exp(w_k\cdot x)},\quad k=1,\cdots,K-1$$

$$P(Y=K|x)=\frac{1}{1+\sum_{k=1}^{K-1}\exp(w_k\cdot x)}$$

- 参数估计方法与二项逻辑斯蒂回归类似

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

- 模型
- 总结

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

算法优点

- 计算代价不高，容易理解实现，在时空表现上相当高效
- 有概率意义，适合需要得到一个分类概率的场景
- 参数代表每个特征对输出的影响，可解释性强

算法缺点

- 容易欠拟合，分类精度不高
- 数据特征有缺失或者特征空间很大时表现效果并不好
- 常作为基准分类器，与其他算法进行比较