浙江大學

《数据建模与分析》Homework

实验题目	Homework 4
学生姓名	莫非
学　　号	3220103345
班　　级	图灵 2201
所在学院	竺可桢学院
提交日期	2024 年 3 月 31 日

1 Homework 3(3.22)

6.2

写出逻辑斯谛回归模型学习的梯度下降算法。

6.2 Solution

直接推广到多项逻辑斯蒂回归。对给定的训练数据集 $T=\{(x_1,y_1),(x_2,y_2),...,(x_N,y_N)\}$ ，其中 $x\in\mathbb{R}^n, y\in\{1,2,...,K\}$ 。为了方便，将权重向量和输入向量加以扩充：

$$w=(w^0,w^1,...,w^n,b)^T, x=(1,x^1,...,x^n,1)^T$$

逻辑斯谛回归模型为：

$$\begin{aligned} P(Y=k|x) &= \frac{\exp(w_k\cdot x)}{1+\sum\limits_{i=1}^{K-1}\exp(w_i\cdot x)}, \quad k=1,2,...,K-1 \\ P(Y=K|x) &= \frac{1}{1+\sum\limits_{i=1}^{K-1}\exp(w_i\cdot x)} \end{aligned} \tag{1.1}$$

对数似然函数为：

$$L(w)=\sum_{i=1}^{N}\sum_{j=1}^{K}I(y_i=j)\log P(Y=j|x_i) \tag{1.2}$$

注意到 $\sum_{j=1}^{K}I(y=j)=1$ ，上式可化简为：

$$L(w)=\sum_{i=1}^{N}\left(\sum_{j=1}^{K-1}I(y_i=j)(w_j\cdot x_i)-\log\left(1+\sum_{j=1}^{K-1}\exp(w_j\cdot x_i)\right)\right) \tag{1.3}$$

Algorithm 1: （多项）逻辑斯谛回归模型学习的梯度下降法

输入：目标函数 $f(w)=-L(w)$ ，梯度函数 $g(w)=\nabla L(w)$ ，计算精度 ε ；
输出：最优参数值 w^* ；最优模型 $P(y|x;w^*)$ 。

1 取初值 $w^{(0)}\in\mathbb{R}^{n+1}$ ，置 $k=0$ 。
2 计算 $f(w^{(k)})$ 。
3 计算梯度 $g(w^{(k)})$ ，当 $\|g_k\|<\varepsilon$ 时，停止迭代，令 $w^*=w^{(k)}$ ；
否则，令 $p_k=-g(w^{(k)})$ ，求 λ_k ，使

$$f(w^{(k)}+\lambda_k p_k)=\min_{\lambda\geq 0}f(w^{(k)}+\lambda p_k)$$

4 置 $w^{(k+1)}=w^{(k)}+\lambda_k p_k$ ，计算 $f(w^{(k+1)})$ 。
当 $|f(w^{(k+1)})-f(w^{(k)})|<\varepsilon$ 或 $\|g(w^{(k+1)})\|<\varepsilon$ 时，停止迭代，令 $w^*=w^{(k+1)}$ ；
5 否则，置 $k=k+1$ ，转第 3 步。

6.3

写出最大熵模型学习的 DFP 算法。(关于一般的 DFP 算法参见附录 B)

6.3 Solution

对最大熵模型而言,

$$P_w(y|x) = \frac{1}{Z_w(x)} \exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right) \tag{1.4}$$

其中规范化因子:

$$Z_w(x) = \sum_y \exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right) \tag{1.5}$$

求解对偶问题的外部极大化问题:

$$\max_w \Psi(w) = \max_w \sum_{x,y} \hat{P}(x,y) w_i f_i(x,y) - \sum_x \hat{P}(x) \log Z_w(x) \tag{1.6}$$

即目标函数为:

$$f(x) = -\Psi(w) = \sum_x \hat{P}(x) \log \sum_y \exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right) - \sum_{x,y} \hat{P}(x,y) w_i f_i(x,y) \tag{1.7}$$

Algorithm 2: 最大熵模型学习的 DFP 算法

输入:特征函数 $f_1, f_2, ..., f_n$;经验分布 $\hat{P}(x,y)$,目标函数 $f(w)$,梯度函数 $g(w) = \nabla f(w)$,计算精度 ε。

输出:最优参数值 w^*;最优模型 $P(y|x;w^*)$。

1 取初值 $w^{(0)} \in \mathbb{R}^n$,取 G_0 为正定对称矩阵,置 $k=0$。

2 计算 $g_k = g(w^{(k)})$,若 $\|g_k\| < \varepsilon$,则停止迭代,令 $w^* = w^{(k)}$;否则,转第 3 步。

3 置 $p_k = -G_k g_k$。

4 一维搜索:求 λ_k,使

$$f(w^{(k)} + \lambda_k p_k) = \min_{\lambda \geq 0} f(w^{(k)} + \lambda p_k)$$

5 置 $w^{(k+1)} = w^{(k)} + \lambda_k p_k$。

6 计算 $g_{k+1} = g(w^{(k+1)})$,若 $\|g_{k+1}\| < \varepsilon$,则停止迭代,令 $w^* = w^{(k+1)}$;否则,按下式求出 G_{k+1}:

$$G_{k+1} = G_k + \frac{\delta_k \delta_k^T}{\delta_k^T y_k} - \frac{G_k y_k y_k^T G_k}{y_k^T G_k y_k}$$

其中,$y_k = g_{k+1} - g_k$,$\delta_k = w^{(k+1)} - w^{(k)}$。

7 置 $k = k+1$,转第 3 步。