机器学习导论习题课

第二次作业

题目1:多标记Logistic回归

在多标记任务中，每个示例都有一个标记集合$\boldsymbol{y} = \{y_1, y_2, \ldots, y_l\}$并且标记满足$y_i \in \{0,1\}$。假设后验概率满足条件独立性：

$$\mathrm{P}(y|x) = \prod_{i=1}^{l} \mathrm{P}(y_i|x)$$

请使用Logistic回归方法解决一下问题。

(1) 请给出你的Logistic回归模型的对数似然函数；

(2) 请计算你对数似然函数的梯度。

题目1:多标记Logistic回归

(1) 解：对于特定标记y_j,其logistic回归模型的对数似然函数可以写成：

$$l(w,b)=\sum_{i=1}^{m}\ln P(y_j^i|x^i;w,b)$$

令$\beta=(w;b)$，$\hat{x}=(x;1)$则，

$$P(y_j^i=1|x^i;w,b)=\frac{e^{y_j^i\beta_j^T\hat{x}^i}}{1+e^{\beta_j^T\hat{x}^i}}$$

$$\ln P(y_j^i|x^i;w,b)=y_j^i\beta_j^T\hat{x}^i-\ln\left(1+e^{\beta_j^T\hat{x}^i}\right)$$

$$l(\beta)=\sum_{i=1}^{m}\left(y_j^i\beta_j^T\hat{x}^i-\ln\left(1+e^{\beta_j^T\hat{x}^i}\right)\right)$$

题目1:多标记Logistic回归

由于后验概率满足条件独立性，所以：

$$\ln P(y|x) = \sum_{j=1}^{l} \ln P(y_j|x)$$

$$l(\beta) = \sum_{j=1}^{l}\sum_{i=1}^{m}(y_j^i \beta_j^T \hat{x}^i - \ln(1 + e^{\beta_j^T \hat{x}^i}))$$

(2) 答：梯度为

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{m}(y_j^i \hat{x}^i - \frac{\hat{x}^i e^{\beta_j^T \hat{x}^i}}{1 + e^{\beta_j^T \hat{x}^i}}) = \sum_{i=1}^{m}(y_j^i \hat{x}^i - \hat{x}^i P(y_j^i = 1|x^i; w, b))$$

$$\nabla_{\beta_\cdot} l(w,b) = [\frac{\partial l(w,b)}{\partial \beta_1}, \dots, \frac{\partial l(w,b)}{\partial \beta_l}]$$

这道题目比较简单。同学们最大问题是上下标标注很不准确，经常会丢掉i或者j。

题目1:多标记Logistic回归

Learning And Mining from DatA
http://lamda.nju.edu.cn

第(2)问的本意只是让大家求到每个标记的梯度就可以了，不过也有人求解了每个标记在所有纬度的梯度，这个答案也被接受。

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{m}\left(y_j^i \hat{x}^i - \frac{\hat{x}^i e^{\beta_j^T \hat{x}^i}}{1+e^{\beta_j^T \hat{x}^i}}\right) = \sum_{i=1}^{m}\left(y_j^i \hat{x}^i - \frac{\hat{x}^i \exp\{\sum_{t=1}^{d}\beta_{jt}\,\hat{x}_t^i\}}{1+\exp\{\sum_{t=1}^{d}\beta_{jt}\,\hat{x}_t^i\}}\right)$$

$$\frac{\partial l(\beta)}{\partial \beta_{jt}} = -\sum_{i=1}^{m}\frac{\hat{x}^i \hat{x}_t^i \exp\{\beta_{jt}\hat{x}_t^i\}(1+\exp\{\sum_{t=1}^{d}\beta_{jt}\,\hat{x}_t^i\}) - \hat{x}^i \exp\{\sum_{t=1}^{d}\beta_{jt}\,\hat{x}_t^i\}\hat{x}_t^i \exp\{\beta_{jt}\hat{x}_t^i}{(1+\exp\{\sum_{t=1}^{d}\beta_{jt}\,\hat{x}_t^i\})^2}$$

题目1:多标记Logistic回归

所以梯度为一个$l \times d$的矩阵，其第(j,t)个元素为：

$$[\nabla_{\beta_{..}} l(w,b)]_{jt} = \frac{\partial l(\beta)}{\partial \beta_{jt}}$$

请注意这个梯度不是一个二阶梯度，而是对一个矩阵变量的一阶梯度，这个梯度的求解更加复杂，而且涉及的参数和符号更多，也就意味着更多的错误。

非常不幸，助教没有发现一个同学算对了这个矩阵梯度，所以计算这个梯度的同学分数可能会更低。

题目2: 线性判别分析

证明对线性回归后得到的$\hat{Y}$做LDA和对原始空间X做LDA结果是等价的。

证明：原始空间下的的类内散度矩阵为：$S_w = \sum_{x \in X_i}(x - u_i)(x - u_i)^\top$，$u_i$为第i个类的类中心。$X_i$表示第i个类的所有样本。类间散度矩阵为：$S_b = \sum_{i=1}^{N} m_i\,(u_i - u)(u_i - u)^\top$，$m_i$表示第i个类的样本个数，u是类中心的中心。

变换后的空间下，

$$S_w^y = \sum_{y \in \hat{Y}_i}\left(y - u_i^y\right)\left(y - u_i^y\right)^\top = \sum_{x \in X_i} B^\top (x - u_i)(x - u_i)^\top B = B^\top S_w B$$

$$S_b^y = \sum_{i=1}^{N} m_i\left(u_i^y - u^y\right)\left(u_i^y - u^y\right)^\top = \sum_{i=1}^{N} m_i\,(B^\top u_i - B^\top u)(B^\top u_i - B^\top u)^\top = B^\top S_b B$$

题目2: 线性判别分析(2)

原始空间下的目标式为：$\mathrm{S_b W} = \lambda \mathrm{S_w W}$

新空间下的目标公式为：$\mathrm{B^\top S_b B W^y} = \lambda^y \mathrm{B^\top S_w B W^y}$

可令：$\mathrm{BW^y} = \mathrm{W}$

则，对$\hat{\mathrm{Y}}$做投影变换可得：$\hat{\mathrm{Y}}\mathrm{W^y} = (\mathrm{XB})(\mathrm{B^{-1}W}) = \mathrm{XW}$。

所以，$\hat{\mathrm{Y}}$上的LDA和X上的LDA等价。

题目3:编程题

题目要求:

1. 实现LR算法，优化算法可选择梯度下降，亦可选择牛顿法
2. 利用“一对多”(One vs. Rest, OvR)策略对分类LR算法进行改进，处理多分类任务

题目3:编程题

一般实验报告需要从 方法-实现-实验结果 三方面来写，在方法中需要简要介绍方法的思路，过程，给出伪代码

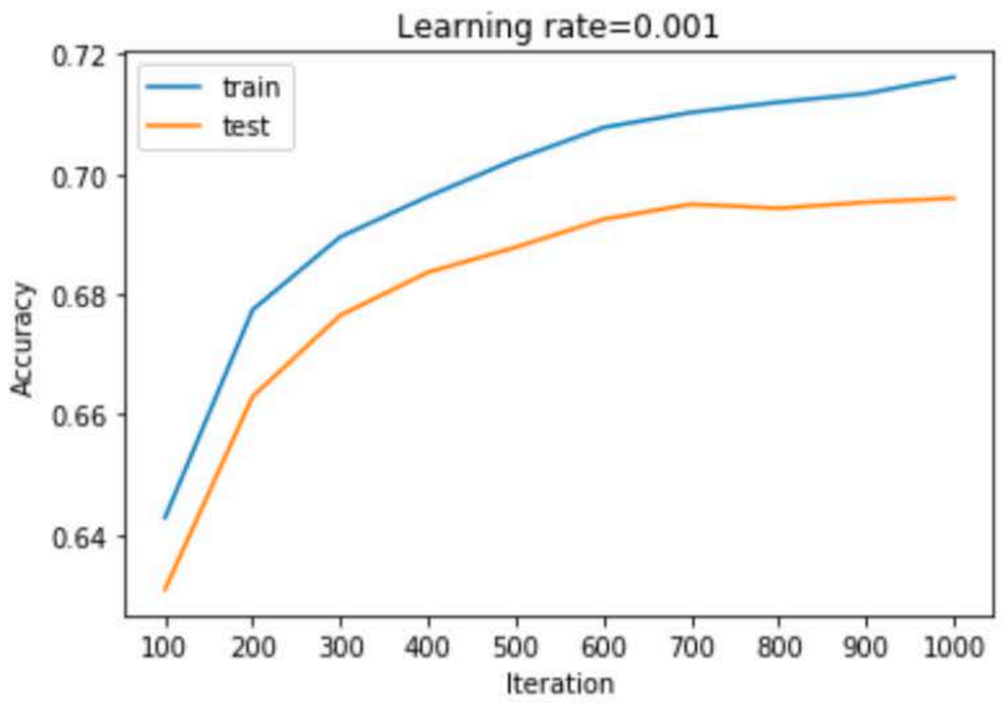
在实现中，需要结合方法给出算法实现的模块划分，显示出自己代码实现的条理性。

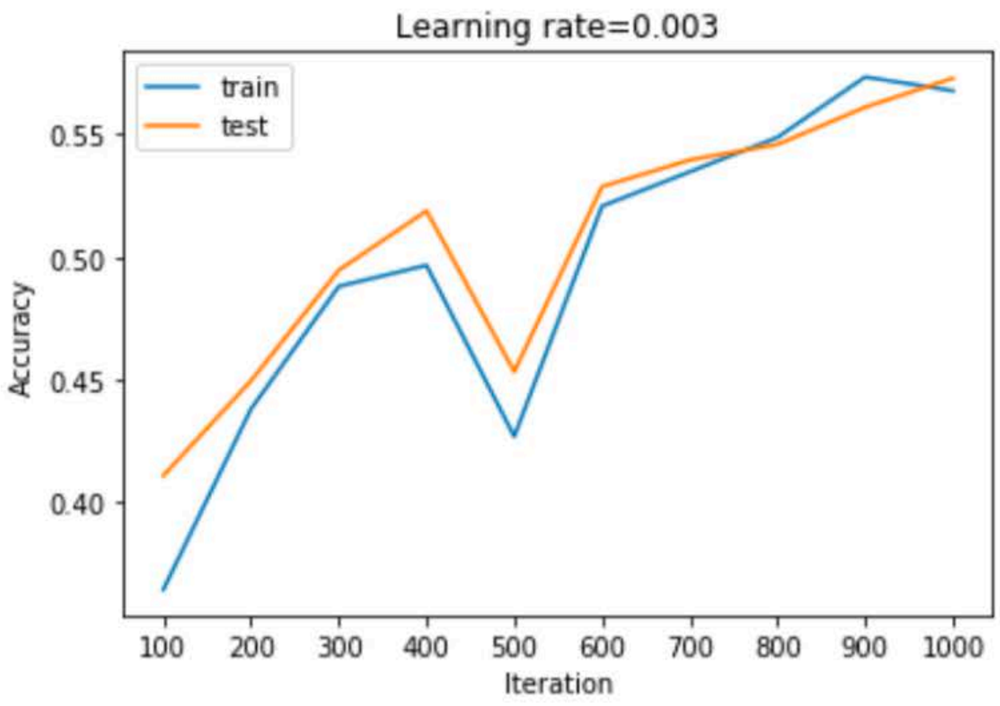
实验结果中，给出题目中要求的实验结果只能说达到了基本要求，方法只有通过对比才能展示其有效性，所以大家要给出实验的对比效果。参数调查也是一项比较重要的内容，如优化方法中的超参设置等。

鼓励大家对实验提出自己的感想或者改进思路。

Effect of Learning Rate

例子





函数功能说明

表 3: 相关函数参数及功能

函数名	参数	功能
sigmod	inX	计算inX对应Sigmoid函数的y值
gradAscent	dataMatIn、classLabels、alpha、maxCycle	对训练数据集dataMatIn和classLabels，设定步长为alpha，迭代次数为maxCycle，用梯度下降法返回对应的w
classifyVector	inX, weights	对每个样本inX，用weights算出对应的预测结果
classifyAll	dataSet, weights	对数据集inX，用weights算出对应的预测结果
Normalize	dataSet	对数据集dataSet进行归一化
getTrainDataSet	无	获得训练数据集
getTestDataSet	无	获得测试数据集
clearResult	trainShares, classNum	将标签trainShares中classNum类的设为0，其他的设为0
Classfy	testDataSet, weights	对测试数据集testDataSet利用26个分类器的weights求得最终的分类预测结果
main	无	开始运行

参考文献

参考文献

[1] Jerome H. Friedman, Robert Tibshirani, Trevor Hastie. The Elements of Statistical Learning. Chapter 4: 106-112.

[2] 统计学习 [The Elements of Statistical Learning] 第四章习题 https://wenku.baidu.com/view/0a8fc43043323968011c927f.html

[3] 机器学习–Logistic 回归计算过程的推导 https://blog.csdn.net/ligang_csdn/article/details/53838743

参考文献的作用：1. 充分尊重前人工作；2. 体现工作量