Lecture 5习题作业

李星毅 U201712072 自实1701

1，讨论OVA和OVO多类分类器算法的各自特点。

解：显然，OVA和OVO都是将之前学过的二分类算法推广应用到多分类问题的策略。

1. OVA多类分类器
   1. 优点：在需要训练的二分类器的个数上看，OVA比OVO效率高，例如对于k分类，只需训练k个二分类器，但对于OVA，则需要个二分类器；每一个二分类问题可以使用PLA、线性回归、Logisitc回归等二分类算法进行分类，其中，基于概率的Logisitc回归的效果比PLA、线性回归的效果更好，因此一般使用类似Logisitc回归的算法作为OVA二分类器；OVA策略还可用于解决多标签分类问题。
   2. 缺点：当类别数比较大的时候，容易出现数据不均衡现象，这将造成分类器具有偏见，分类效果不佳。一个值得一提的workaround是提升数量少的样本的权重，减小数量大的样本的权重，这样可以缓解数据不均衡导致的问题。
2. OVO多类分类器
   1. 优点：在训练一个二分类器时所使用的样本个数上看，OVO比OVA效率可能更高，原因在于OVO每次只考虑两个类别的样本，而OVA每次都考虑所有的训练样本；比较稳定，不易出现数据不均衡现象，分类结果更准确；每一个二分类问题可以使用PLA、线性回归、Logisitc回归等二分类算法进行分类。
   2. 缺点：OVO 用时较多，占用更多内存空间，预测更慢，需要训练更多的二分类器。

2，现有四个样本，假设样本（3，0）和（3，6）属于第一类，样本（0，3）属于第二类，样本（-3，0）属于第三类，请用Softmax算法设计出这三个类别的分类器（需要写出计算过程）。

解：首先推导一般情况下的Softmax Regression算法。对于样本集合，设样本矩阵为，对应的标签为，类别数为，权重矩阵为

1. 基于Softmax函数，我们的目标是找到一个函数



使其满足

1. 根据交叉熵的定义，我们可以知道损失函数应该定义为



1. 对损失函数求梯度可得：



1. 为了便于编程实现，利用独热编码的概念，假定标签矩阵为



且记



故有



现将上述损失函数梯度表达式转换成矩阵形式，即：



基于此，我们可以写出Softmax Regression算法的步骤如下：

| **Algorithm** **1** Softmax Regression with Batch Gradient Descent |
| --- |
| Initialize  For   1. Compute      1. Update by     until  or enough iterations  return |

| **Algorithm** **2** Softmax Regression with Stochastic Gradient Descent |
| --- |
| Initialize  For  For   1. Compute      1. Update by     until  or enough iterations  return |

1. 值得注意的是，在实际编程时，由于Softmax函数要取指数，并且由数值分析的知识可知，计算机可能出现上下溢出的问题，一个可行的解决办法是利用如下等式，在原先的基础上减去最大值来代替Softmax函数：



1. 可以看出，Softmax Regression算法是Logistic Regression算法的推广，用于多分类问题。与OVO，OVA策略不同的是，Softmax Regression是直接进行执行多分类。而OVO，OVA是把多个二分类器组合到一起完成多分类任务。
2. 因此，现在将推导得到的Softmax Regression算法应用到题干所给的数据集中：

由题意得：，，转化成独热编码为，假设初始化的权重矩阵为，选取学习率。

1. 此时预测标签为

计算损失函数梯度：



更新权重矩阵：



1. 类似前面步骤，可计算出：



1. 当迭代到大约7~8轮时，有



所对应的权值矩阵为：



综上所述，所求的为：



其中，且分类器得出的该样本分类结果为这个概率向量的最大分量所对应的索引。

3，假设有一个样本有d维特征，现选择二次多项式对其进行特征变换，请计算出变换后的特征维数（包含常数项）。

解：由题意知：



可以计算出特征维数应该为。