教程 4 : Softmax

提交日期：2020.8.1

提交人：詹紫琦

目录

[1. 题目 3](#_Toc32066)

[1.1 Softmax回归 3](#_Toc21533)

[1.2 运行环境 3](#_Toc20132)

[2. 算法阐述或实验步骤说明 3](#_Toc3449)

[2.1变量说明 3](#_Toc683)

[2.2 Softmax回归阐述及步骤说明 3](#_Toc5939)

[3. 实验结果与截图 4](#_Toc8626)

[3.1 Softmax回归分类实验结果 4](#_Toc31910)

[4. 总结 8](#_Toc17510)

[4.1 总结 8](#_Toc19328)

[5. 参考文献 8](#_Toc8848)

# 题目

## 1.1 Softmax回归

Softmax[1]是一种比较简单以及传统的回归方法，这种方法虽然称之为回归算法，但与逻辑回归相同的是，这种算法一般都应用于分类算法中。而逻辑回归就是Softmax回归的一种简单形式。Softmax回归名字的由来是由于在预测分类过程中使用了Softmax函数，而该函数其实就是该函数内元素的指数与所有元素指数和的比值。这样，通过Softmax函数就可以将多分类的输出值转换为范围在和为1的概率分布。通过找出最大的概率值就能得到分类结果

## 1.2 运行环境

系统：Ubuntu20.04，python3.7，Anaconda集成工具Jupyter编写。

# 算法阐述或实验步骤说明

## 2.1变量说明

是训练数据集，包含着个类别。

是输入训练集中的类别集合，表示输入数据归属于类别j的概率，是不同参数权重下的代价函数。

是参数矩阵。

## 2.2 Softmax回归阐述及步骤说明

首先读取数据集，将训练集中的属性数据与标签数据分开，根据数据集找出所需分类的类别个数。然后将对应的不同类别的标签数据进行独热编码，假如存在两个类别，则不同类别的独热编码后的标签项为或者是。即只有对应类别的一项为1，其余全为0。进行完数据的处理之后。开始编写softmax函数，通过公式2-1得到每一个数据项对应类别的概率。

 （2-1）

然后进行初始化权重矩阵，通过公式2-1得到了每一个类别的概率，此时开始计算损失函数的大小，损失（代价）函数的计算公式如2-2：

 （2-2）

根据初始化的权重矩阵以及softmax函数以及独热编码后的标签矩阵能得到初始的代价函数的值。由于考虑到可能会出现过拟合现象，所以引入正则项，如代价函数计算公式更新为公式2-3：

 （2-3）

得到代价函数的计算公式后，需要达到最小化代价函数的优化目标，通过更新权重矩阵来实现最优化目标。这里使用的是梯度下降最优化方法。让代价函数对权重矩阵求偏导找出梯度的计算公式，如公式2-4：

 （2-4）

计算出权重参数矩阵的梯度后，设置迭代次数iteration以及学习率alpha进行梯度下降更新，如公式2-5：

 （2-5）

最终能实现最小化代价函数，并且得到相应的权重参数矩阵。

# 实验结果与截图

## 3.1 Softmax回归分类实验结果

首先读取数据集，这里用的是鸢尾花数据集，初步分析数据类别，分为三类分

别用0，1，2替代。得到标签值如下图3-1

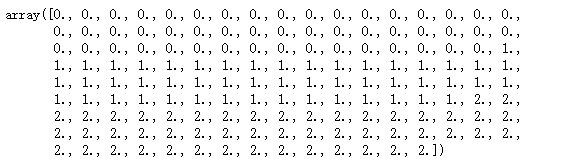


图3-1

进行独热编码后得到的新矩阵前面部分如下：

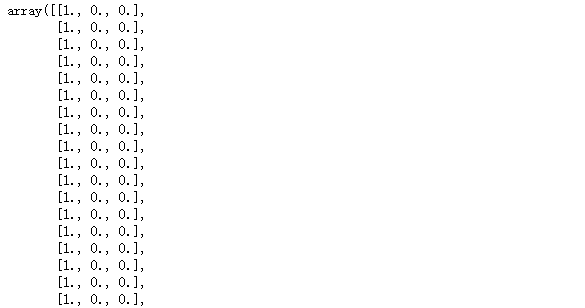


图3-2

进行完独热编码后，定义softmax函数，初始化定义权重矩阵，设置迭代次数为1000，学习率为0.01，正则化参数lamda为0.01，得到的损失函数随迭代次数变化的图像如图3-3：

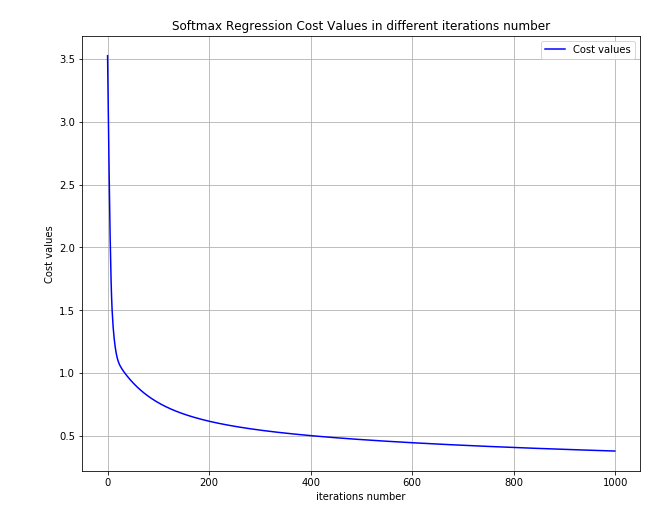


图3-3 损失函数随迭代次数变化的图像，iteration为1000，学习率为0.01

通过图像我们可以看出在这两类参数下损失函数的变化较为平滑，而且下降趋势一直没有变化，符合梯度下降损失函数值的变化情况。

进行完梯度下降后，开始考虑参数迭代次数iteration，学习率alpha对于模型预测准确率的影响。选取不同迭代次数100，500，2500，5000，10000，20000，不同学习率0.1，0.03，0.01，0.003，0.001，0.0003总共包含36个组合。分别计算这36个组合下的分类准确率，得到的结果如图3-4：

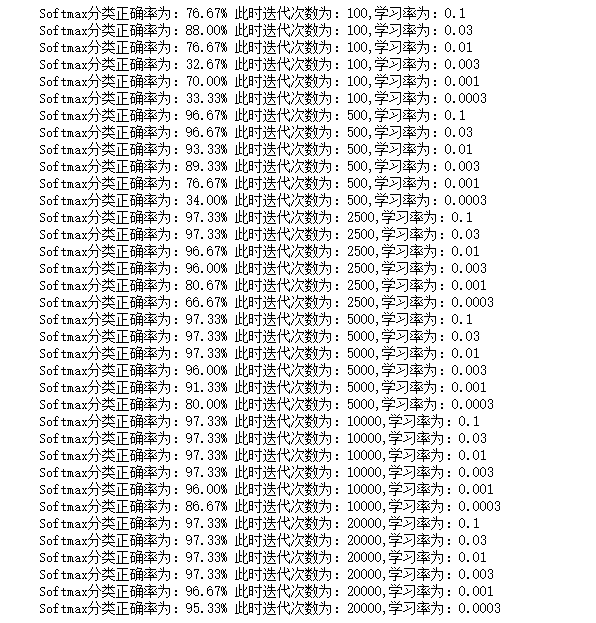


图3-4 36个参数组合下训练的softmax分类正确率

通过训练结果可以看出当迭代次数达到2500时就能达到最高分类准确率。得到的分类正确率的散点图如图3-5：

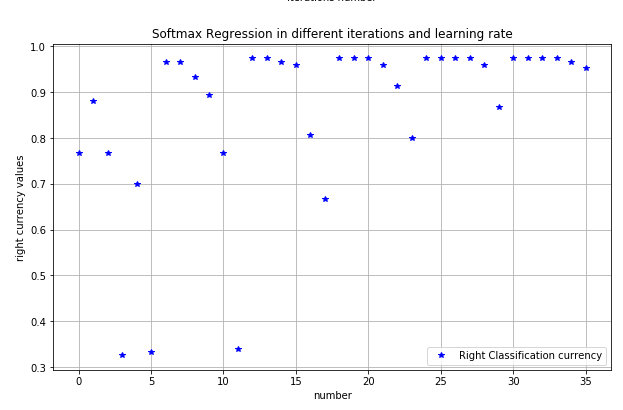


图3-5 6x6=36个参数组合下的分类正确率

# 总结

## 4.1 总结

Softmax函数就是将多分类的输出值转换为范围在和为1的概率分布，引入Softmax函数的优点是能使特征变化明显，斜率能明显增大，这种函数曲线能够将输出的数值拉开距离，经过Softmax函数能够将差距大的数值距离拉的更大。同时在函数求导时也比较方便，由于Softmax引入指数函数，求导较为容易。但是带来的缺点也比较明显，当输入的值非常大时，再通过指数函数，可能会产生数值溢出的现象。

# 参考文献

[1] 周志华等.机器学习（西瓜书）