**Softmax回归——多元分类**

**摘要**

图像分类是把图像中不同目标区分出来的一种图像方法，是计算机视觉判读的重要手段。基于深度学习的图像分类方法取得了突破性进展，并广泛运用于人脸识别，卫星遥感，医疗诊断，自主导航以及人机交互等领域。

对于离散值而非连续值的预测，softmax回归(Softmax Regression)模型能让我们在识别多种分类中的一个，起到了学习已知数据，预测未知数的作用。Softmax回归的输出单元由一个变成了多个，且引入了softmax运算使输出更适合离散值的预测和训练。

本文根据给出的数据集，划分训练集与交叉验证集，用softmax回归对训练集进行学习，得到各个样本在不同分类模型下的占比。接着，使用交叉熵作损失函数进行参数优化。得到的数据后在交叉验证集上进行验证。

1. **问题分析**

由训练数据集可知，每个样本有28\*28个特征，即一张图片有784个像素点，每个像素点的值在0-255之间，每个像素点设为一个x，即一个特征。所以一个样本的特征是一个28\*28长度的向量，每一个特征x在计算过程中都有一个权重，引入权重w建立直线结构。

接着使用softmax计算一组数值中每个的占比，得到一个概率分布，第一个样本属于第一个分类的概率；第一个样本属于第二个分类的概率，依此类推。

计算损失函数，将每个样本对应的label用独热编码表示——真实分类对应的位置对应的概率是1，其他是0。使用梯度下降法，不断更新权重，交叉熵越小，说明损失越少，预测越准确。

**二、符号及符号说明**

|  |  |
| --- | --- |
| **符号** | **符号说明** |
|  |  |
| O  N  M  X  W  P  Q      Z  G | 样本数量  特征数量  类别个数  样本矩阵(o,n)  权重矩阵(m,n)  标签矩阵(o,m)  Softmax计算得到的矩阵(o,m)  单位矩阵—交叉熵计算（o,1）  单位矩阵—softmax分母求和（m,1）  直线结构计算值  对Z指数运算所得值 |

1. **模型建立与求解**

**3.1直线结构**

,n为特征数量

**3.2** **softmax计算**

对应矩阵运算：

设有o个样本，每个样本有n个特征，令X(o,n)表示样本矩阵，每行是一个样本；设有m个类别，W(m,n)表示权重矩阵，每行是一个类别

计算得到Z(o,m)，E1(m,1)是一个单位矩阵，作用是求和运算；得到的Q为(o,m)矩阵。

当softmax的输入比较大的时候，可能会产生overflow，超过了float能表示的范围。同理，当输入的数为负数且绝对值很大的时候，会出现分子分母变得非常小，四舍五入就会为0，导致underflow。

实验过程中，数据陡峭，开始是上溢出，尝试减去最大值之后，又是下溢出。

模型蒸馏文章中提到设计软标签的损失函数时使用了一种特殊的softmax，

T是温度超参数，提出温度值的作者认为不同类别之间的关系不应该知识非0即1的关系，（如：将猫误判成够的损失直观上比将猫误判成汽车的损失小），因此，粗暴地使用one-hot编码丢失了类间和类内的额外信息[1]。T>1，分布均匀化；T<1，分布极端化。

虽然，此次实验没有要区分的特殊意义，但是将数据进行缩放之后，计算不再溢出了。

**3.3交叉熵计算**

交叉熵公式：

Pi是样本的真实分类，qi是softmax计算的结果

矩阵表示：

P(m,1)，Q(o,m)，此处因为向量化的第一个标签，即第一个样本的真实标签应该乘以第一个样本在各个类别模型下的softmax值，同理第二…到最后一个样本，所以要取对角线值组成一个(1,o)的向量，E2是一个(o,1)的单位向量，最后除以样本数，就是整体的损失函数计算方法。

**3.4梯度下降**

损失函数的目标是获得最小值，从正无穷方向无限接近0，通过导数不断更新w的值，求偏导函数：

为学习率，表示每一步更新的步长。

求偏导的结果——真实分布与预测分布的插值乘上权重参数对应的特征值

当预测分布与实际分布越接近时，更新量会变得很小。

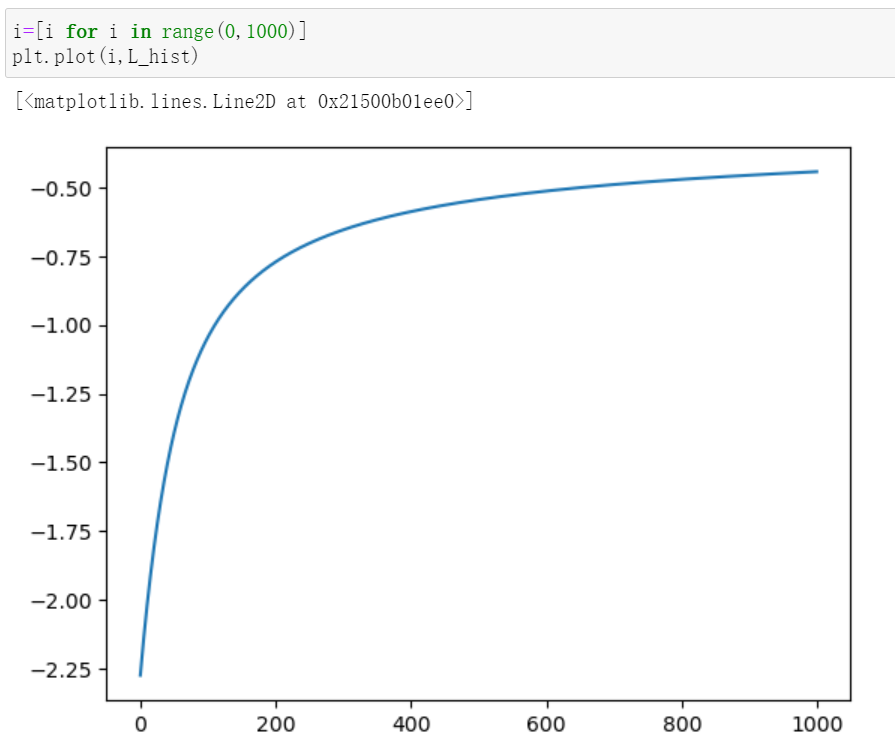
矩阵表示：

P(o,m),Q(o,m),X(o,n),得到D(o,n)的矩阵

学习率是一个敏感的参数，设置太大可能出现不收敛的情况，太小会出现收敛太慢的情况。学习率不同，以及w初始值设置不同会获得不同的效果。

**四、模型评估**

在训练过程中记录下损失函数的变化，下图展示了执行优化过程中的loss变化趋势。Argmax()是将预测结果和真实结果的一维数组中的最大值索引分别提取出来，将预测结果和正确结果进行比较，计算错误率。



**五、结束语**

随着互联网技术的迅速发展及计算机能力的大幅度提升，图像数据的规模不断壮大，对分类算法得到性能要求不断提高。产生了很多softmax的变体，本文采用的模型简单识别率不算很高，其研究和改进对提高图像分类具有重要意义。

**参考文献**

1. 刘冰,李瑞麟,封举富. 深度度量学习综述[J]. 智能系统学报,2019,14(6):1064-1072. DOI:10.11992/tis.201906045.
2. 李林峰,陈佳怡,郑佳伟,等. 基于Tensorflow框架的手写数字识别[J]. 数字通信世界,2023(2):133-136. DOI:10.3969/J.ISSN.1672-7274.2023.02.044.