**《统计学习理论及应用》实验二（感知机模型）报告**

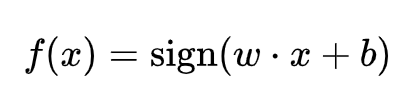
**刘洋岑 2020080601018**

**一. 问题描述**

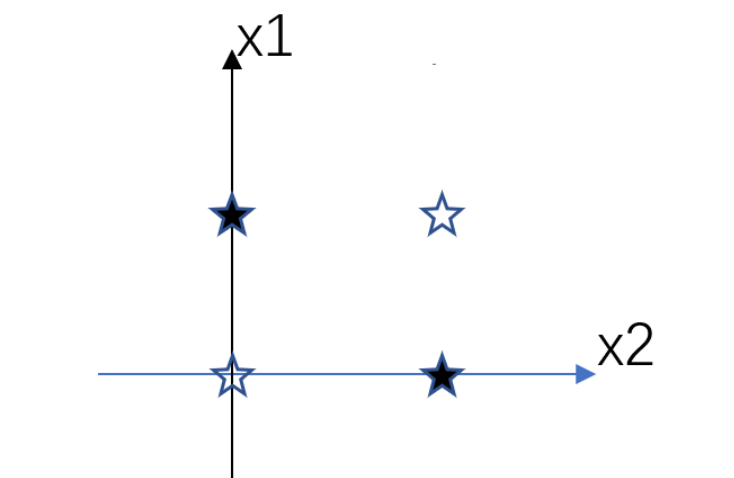
使用感知机模型进行Mnist的分类任务，与或门的实现任务。

**二. 解决方案模型描述和原理图**

单层感知机采用的是线性函数作为内核，外部使用一个激活函数进行激活。而多层感知机则是将多个单层感知机进行联合，形成多个层次进行传播。



这种形式上与全连接神经网络似乎是相同的。我认为，其区别主要在于以下几点：

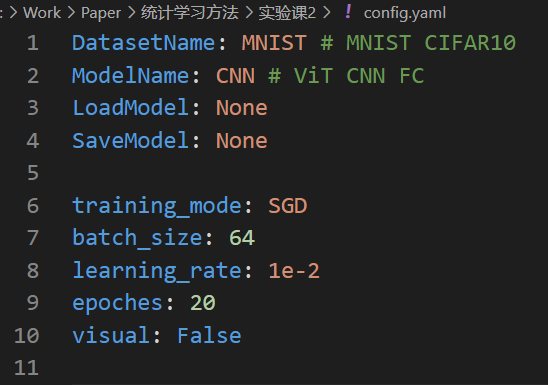


比如上图展示了一个期望用单层感知机实现的异或门，使用普通的线性模型是无法做到的。但是加入了感知层中的激活函数，可视化进行观测是可以实现“曲线”的。

**三. 算法、公式和流程图**

代码放置在了github，附链接：

<https://github.com/Randle-Github/statistical_learning> 中的实验二部分。由于《统计学习方法》课程设置和《机器学习综合实验》重合了，我直接把那边的代码clone过来。我一共实现了两种数据集以及三种模型，在config.yaml中进行设置。本课程要求的多层感知机则将模型选择为FC。



在PPT中使用了对偶优化的办法，我这边是纯的梯度下降方法。不过我理解的对偶优化方法反而是比直接用损失函数更加基础的方法。必须将多分类问题转化为2分类问题来做。

1. **计算指标，结果图、表和分析**

计算指标使用的简单的top-1 error：



CNN(demo)是机器学习实验课老师给的代码。其它的是我自己写的，至于ViT，我使用的单head的ViT结构，本来SOTA也是用ViT实现的，结果我的代码里面一直不传梯度，可能是在某些拼接的地方把梯度丢失了，所以实验失败。

可以看出CNN确实具备一定的提取图像特征的能力。使用单纯的FC效果差距还是比较显著。

用线性感知机做与或门的实验我就没有亲手操作了，因为这个基本从公式就可以看出加入激活函数后是可以很好的拟合的。

对于神经网络的“拟合任意函数”有一种解释的理论就是从与或门的角度进行解释，对于多层感知机，前面的部分层如果形成了“与或门”的结构，后续是可以任意拟合函数的。

**五. 方案改进计划**

对于多层感知器，是一个非常基础的模型。解决多分类问题时有很大的局限性。假设分类过多，需要对于整体进行一个划分，容易造成类别不平衡的问题。这时损失函数的设计就变得极为重要了。在感知机中没有这一概念的存在，所以这方面也是后续的改进。