人工智能基础 实验二

PB18151866 龚小航

**【实验目的】**

本次实验包含传统机器学习与深度学习两部分。通过本次实验初步了解机器学习的基本方法。

**【实验内容】**

在传统机器学习部分实现线性分类算法、朴素贝叶斯分类器以及SVM算法；在深度学习部分实现手写感知机模型并进行反向传播与复现MLP-Mixer。

传统机器学习部分需要根据鲍鱼的物理测量属性预测鲍鱼的年龄，label已经经过处理分为1，2，3三类，用各种算法在测试集上预测鲍鱼的年龄组别，并分析预测准确度。

深度学习部分首先需要实现四层感知机模型，神经元个数为5，4，4，3，激活函数选为sigmoid。在此基础上实现误差逆传播算法与梯度下降算法，使用随机输入与随机标签。而MLP-Mixer部分需要复现MLP-Mixer模型，实现手写数字图像分类识别。

**【实验过程与运行结果】**

一 、实现线性分类器

对鲍鱼数据集进行线性分类。鲍鱼数据集对一条训练数据共有8个属性和一个类别标签。线性分类器需要找到一个使得所有训练数据到分割线的最小二乘距离最小。引入L2规范化项之后，线性分类问题等价于求解以下优化问题：

直接对求偏导并令其 ：

得到这个优化问题的解，因此只需要令取上述值即可完成最小二乘线性分类。

需要注意的是，在计算中已经把参数 吸收入向量形式，相应的，把训练集再在最右侧加上一列全1。因此上述 写成矩阵形式时具有如下形式：

更新参数的过程如下所示，输入训练集，返回最小二乘分类参数。

图形用户界面, 文本, 应用程序

描述已自动生成

再利用得到的参数在测试集上作预测，输入测试数据集（无label），返回对所有测试数据的预测类别：

图形用户界面, 文本, 应用程序

描述已自动生成

最后运行统计正确率：

图形用户界面, 文本, 应用程序

描述已自动生成

二、实现朴素贝叶斯分类器

朴素贝叶斯分类器主要需要实现每个类别的概率以及条件概率和先验概率。首先是计算每个label出现的概率，此处使用拉普拉斯平滑来处理没有在训练集中出现过的label的概率。计算式如下所示：

即训练集D中，类别（Label）为c出现的概率为训练集中类别为c的数据的条目数+1除以训练集中所有的数据条目数|D|+Label的总类别数|N|。分母上加|N|是为了平衡分子上为了平滑而加的1

而对于先验概率，对于连续型和离散性数据需要分别处理。

对于离散型属性，利用下式计算先验概率：

其中表示类别为c，且第i个属性值为的数据条目数，表示第个属性可能的取值数。

对于连续型数据，可以认为它服从正态分布，直接从训练集中得出这一条属性的 即可完全确定它的分布情况。需要注意的是，这一步需要根据label类别分组计算。比如训练集包含10种label，那么对于一个 连续型属性就要有10组值。

正态分布的函数定义在全局，输入返回计算结果：

文本

中度可信度描述已自动生成

有了所有的概率数据之后，就可以在测试集上预测分类结果。对于一条测试数据而言，判断它属于哪一类通过下式进行。对于连续型数据，累乘项直接用分布函数一步计算到位。

最后的运行结果与正确率如下所示：

图形用户界面, 文本, 应用程序

描述已自动生成

三、实现SVM分类器

要求完成支持软间隔和核函数的SVM。

需要满足的对偶问题可以写为：（已经化为标准形式）

调用cvxopt模块引入二次规划求解包，qp方法要求的参数（P,Q,G,H,A,B）形式如下：

按规则写出这六个matrix格式的矩阵。只需要一一对应即可。

最后求解出，按以下方式计算与. 的维度与一条数据(x)的维度相同,

得出以上结果后，再对测试集直接进行 即可得到预测分数

最终得到的运行结果如下所示：

图形用户界面, 文本, 应用程序

描述已自动生成

四、手写感知机模型并进行反向传播

本部分通过定义MLP模型并将其例化的方式实现。

首先定义隐藏层类HiddenLayer和回归层类LogisticRegression. 类中有属性。根据以下公式在每一条训练数据读入时更新参数矩阵：

文本, 信件

描述已自动生成

将所有训练数据都过一次称为一轮迭代。

经过1000轮迭代，损失曲线如下所示：

图形用户界面, 文本, 应用程序

描述已自动生成

调用自动求导模块得到的结果，可以看出自动求导收敛速度较快。

图形用户界面, 应用程序

描述已自动生成

五、复现MLP-Mixer

此部分可以调用torch的所有功能，只需要在待补区域按说明填入实现代码即可。整体结构如下所示：

图示

描述已自动生成