

《大数据分析方法》课程实验报告

学号： 2020204246

姓名： 王浩

专业： 计算机科学与技术

班级： 20 图灵

实验四 感知器算法的设计实现

实验目的：

1. 熟悉感知器算法。
2. 掌握感知准则函数分类器设计方法。
3. 掌握感知器算法，利用它对输入的数据进行多类分类

实验内容：

1. 数据生成及规范化处理，利用高斯模型，生成 N 类 ($N > 5$) 数据 (2D or 3D)，并对生成样本进行规范化处理
2. 基于生成数据，利用感知器准则实现多类分类，得到分界面的表达式。
3. 生成测试数据列，并对测试数据进行分类判别

实验要求： 给出数据处理过程说明，算法预测过程，结果说明

一. 问题描述

首先生成满足高斯分布的类别数据，通过数学公式建立感知器模型，然后模型训练数据，最终求出模型参数，并将分类结果可视化

二. 解决思路

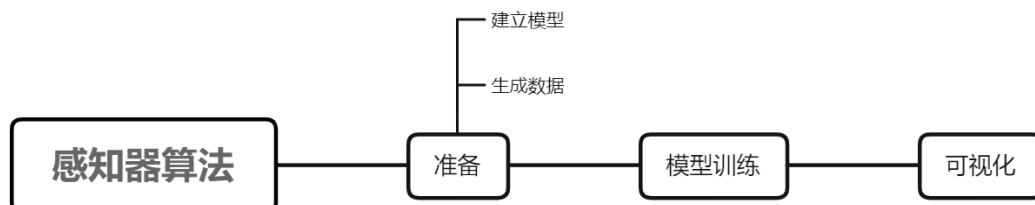


图 1 问题解决流程图

2.1 数据集处理

2.1.1 生成数据

利用`np.random.multivariate_normal`函数生成满足高斯分布的数据并添加对应标签。

```
X.append(np.random.multivariate_normal(mean, cov, size))  
y.extend([i]*size)
```



图 2 生成数据可视化

2.1.2 模型建立

3. 利用感知器实现多类判别

基本算法如下：

- (1) 样本规范化
- (2) 每一类设定一个初始权向量 $w_i(i = 1, \dots, M)$
- (3) 对第 i 类的样本 y_j , 若

$$w_i^T y_j \leq w_t^T y_j \quad t = 1, \dots, M, \text{但 } t \neq i$$

则：

$$w_i(k+1) = w_i(k) + y_j$$

$$w_t(k+1) = w_t(k) - y_j \quad t \neq i$$

- (4) 对所有样本重复 (3)，直到满足

$$w_i^T y_j > w_t^T y_j \quad t = 1, \dots, M, \text{但 } t \neq i$$

图 3 算法原理图

利用上述数学算法原理，进行感知器模型的设计

```
def Perception(X_train, y_train, max_iter_num = 1000):
    X_train = Homogenization(X_train)
    kindNum = len(np.unique(y_train))
    w = np.zeros((kindNum, 3))

    isConverge = False
    while(not isConverge and max_iter_num > 0):
        isConverge = True
        for i in range(len(y_train)):
            x = X_train[i]
            y = y_train[i]
            for j in range(kindNum):
                if i == j:
                    continue
                if np.dot(w[y], x) <= np.dot(w[j], x):
                    isConverge = False
                    w[y] = w[y] + x
                    w[j] = w[j] - x
            max_iter_num = max_iter_num - 1
    return w
```

三. 结果分析

3.1 模型结果

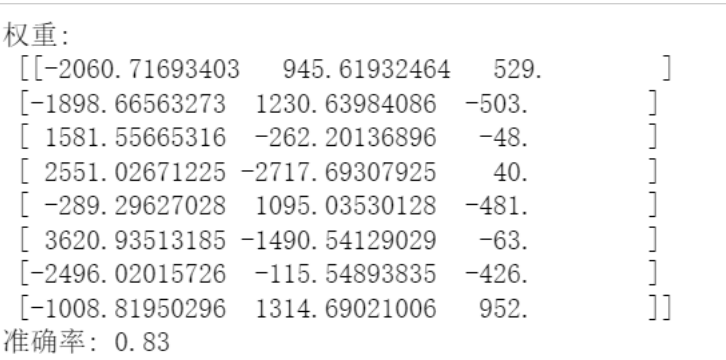


图 4 权重和准确率图

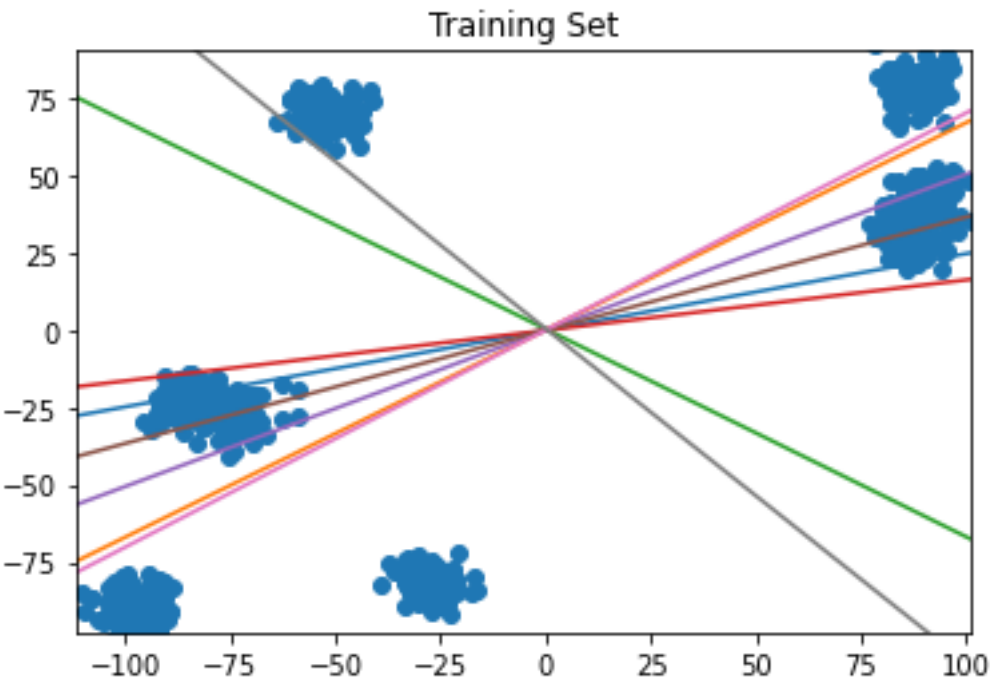


图 5 分类线图

3.2 实验中存在的问题及解决方法

首先由于数据是随机生成的，对最后模型精度有一定的影响，导致精度在80%到90%之间变化。

其次，一开始将数学公式转化为代码的时候，矩阵乘法，没有用dot函数，而是直接用'*'符号来写代码，导致结果出了问题，后来改用dot实现矩阵乘法。