**北京邮电大学软件学院**

**2018－2019学年第一学期实验报告**

**课程名称： 云计算数据中心**

**项目名称： 机器学习基础**

**项目完成人：**

**姓名：\_\_\_田瑾\_\_\_\_\_学号：\_\_\_\_2016522038\_\_\_\_**

**指导教师： \_\_管皓\_\_\_\_\_\_\_\_\_\_\_**

**日 期： 201 年 1 月 10 日**

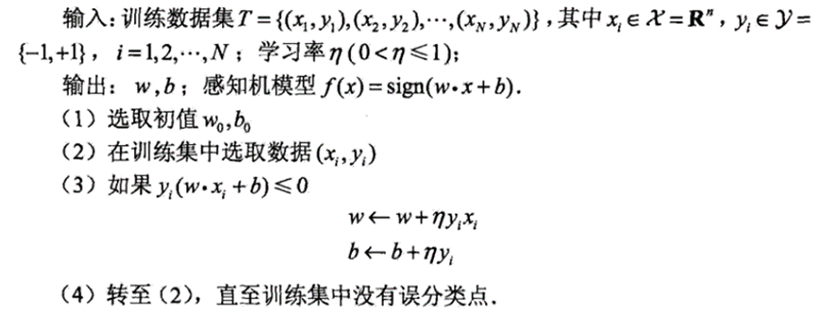
1. **实验目的**

学生通过编程，初步掌握机器学习中的感知机算法原理，为进一步研究更高级的智能算法（神经网络与深度学习）打下基础。

1. **实验内容**

概述：

感知机算法是近代神经网络的重要基础，它是一种二类分类的线性分类模型。其基础的算法流程如下：



题目：

现有二维空间的样本点 (x,y),样本分为两类(类别1，类别-1)。现提供了20个数据样本及其标签供使用（train\_data）。有5个未知样本待分类(test\_data)。

(1) 请可视化训练样本点。

(2) 编程实现感知机算法。

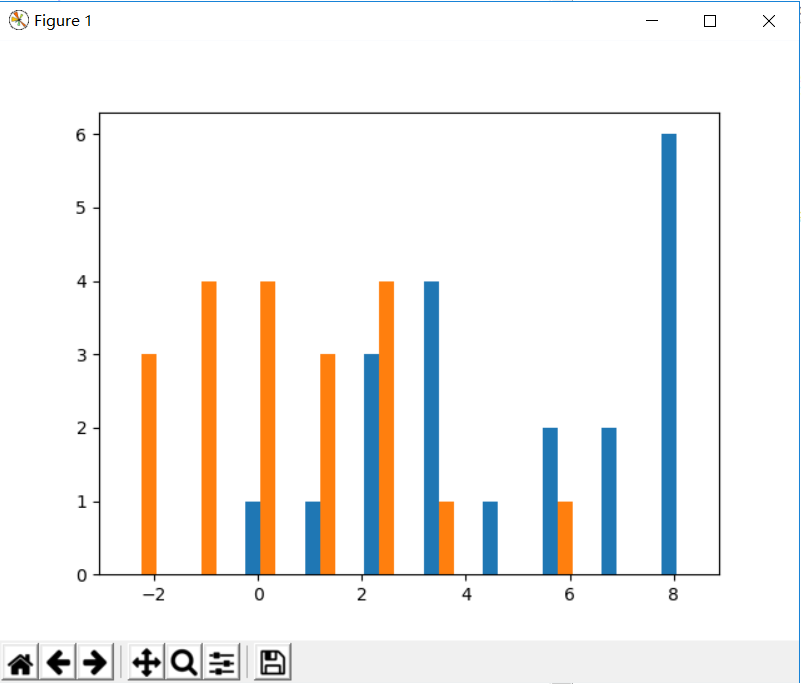
(3) 可视化训练完成的感知机。

(4) 对5个未知样本点进行分类。

1. **实验环境**

Windows操作系统

1. **实验结果**



使用python语言完成实验

首先将excel文件导出.csv文件，方便使用numpy.loadtxt读取数据，画图使用的matplotlib.pyplot库，画训练集直方图。

代码：

import matplotlib.pyplot as plt

import numpy as np

np.set\_printoptions(suppress=True)

traindata = np.loadtxt(open("D:/train.csv","rb"),delimiter=',',skiprows=0)

print(traindata)

testdata = np.loadtxt(open("D:/test.csv","rb"),delimiter=',',skiprows=0)

print(testdata)

k = traindata[:,:2]

plt.hist(k,rwidth=0.5)

#plt.show()

t=traindata[1:]

print(t)

class Perceptron(object):

def \_\_init\_\_(self,eta=0.01,n\_iter=10):

self.eta = eta

self.n\_iter = n\_iter

def fit(self,X,y):

self.w\_ = np.zeros(1 + X.shape[1])#先对权重参数初始化，然后对训练集中每一个样本循环，对权重进行更新。

self.errors\_ = []#记录每轮误分样本

for \_ in range(self.n\_iter):

errors = 0

for xi, target in zip(X,y):#关联样本特征和预测

update = self.eta \* (target - self.predict(xi))

self.w\_[1:] += update \* xi

self.w\_[0] += update

errors += int(update != 0.0)

self.errors\_.append(errors)

return self

def net\_input(self,X):

return np.dot(X,self.w\_[1:]) + self.w\_[0]

def predict(self,X):#预测类别标记

return np.where(self.net\_input(X) >= 0.0,1,-1)

ppn = Perceptron(eta = 0.1,n\_iter = 10)

ppn.fit(testdata,traindata)

plt.plot(range(1,len(ppn.errors\_)+1),ppn.errors\_,marker='o')

plt.xlabel('Epoches')

plt.ylabel('Number of misclassifications')

plt.show()

1. **附录**

在读取test数据的时候，总是显示格式不对打不开，使用Excel重新导出也不行，于是就用txt打开手动更改格式，然后将扩展名改为csv最终解决问题。