point={'0':0,'1':1, '2':2 ,'3':3, '4':4, '5':5, '6':6, '7':7, '8':8,'9':9, '10':10, '11':11, '12':12, '15':1, '30':2, '40':3, 'AD':4}

## 数据预处理

import csv

filename='Wimbledon\_featured\_matches.csv'

l=[]

with open(filename,'r',encoding="utf-8") as csvfile:

    reader = csv.DictReader(csvfile)

    for row in reader:

        # 计算变量值

        point\_Faqiu1 = (1 if int(row['server'])==1 and int(row['point\_victor'])==1 else 0) + int(row['p1\_ace']) + int(row['p1\_winner'])

        point\_Faqiu2 = (1 if int(row['server'])==2 and int(row['point\_victor'])==2 else 0) + int(row['p2\_ace']) + int(row['p2\_winner'])

        # point\_Faqiu1 = (1 if int(row['server'])==1 and int(row['point\_victor'])==1 else 0)

        # point\_Faqiu2 = (1 if int(row['server'])==2 and int(row['point\_victor'])==2 else 0)

        point\_Jieqiu1 = 1 if int(row['server'])==2 and int(row['point\_victor'])==1 else 0

        point\_Jieqiu2 = 1 if int(row['server'])==1 and int(row['point\_victor'])==2 else 0

        # d1 = float(row['p1\_distance\_run'])

        # d2 = float(row['p2\_distance\_run'])

        point\_Error1 = int(row['p1\_double\_fault']) + (1 if int(row['server'])==1 and int(row['serve\_no'])==2 else 0)

        point\_Error2 = int(row['p2\_double\_fault']) + (1 if int(row['server'])==2 and int(row['serve\_no'])==2 else 0)

        point\_delta1 = int(point[row['p1\_score']]-point[row['p2\_score']])

        point\_delta2 = int(point[row['p2\_score']]-point[row['p1\_score']])

        # 保存到列表

        l.append([int(row['match\_id'][-4:]), point\_Faqiu1, point\_Faqiu2,

                  point\_Jieqiu1, point\_Jieqiu2,

                #   d1, d2,

                  # 0.0,0.0,

                  point\_Error1, point\_Error2,

                  point\_delta1, point\_delta2,

                  int(row['point\_victor'])])

print(l)

## 数据标准化

## x= [i-min(x) / max(x)-min(x) for i in x]

import numpy as np

# 比赛号集合

matchid = list(set(np.array([i[0] for i in l])))

matchid.sort()

# print(matchid)

last\_id = matchid[0]

index1 = []

index2 = []

index3 = []

index4 = []

index5 = []

for item in l:

    if item[0] == last\_id:

        if index1 == []:

            index1.append(item[1])

            index1.append(item[2])

            index2.append(item[3])

            index2.append(item[4])

            index3.append(item[5])

            index3.append(item[6])

            index4.append(item[7])

            index4.append(item[8])

            # index5.append(item[9])

            # index5.append(item[10])

        else:

            # 累积

            index1.append(index1[-2]+item[1])

            index1.append(index1[-2]+item[2])

            index2.append(index2[-2]+item[3])

            index2.append(index2[-2]+item[4])

            index3.append(index3[-2]+item[5])

            index3.append(index3[-2]+item[6])

            index4.append(index4[-2]+item[7])

            index4.append(index4[-2]+item[8])

            # index5.append(index5[-2]+item[9])

            # index5.append(index5[-2]+item[10])

    else:

        last\_id = item[0]

        index1.append(item[1])

        index1.append(item[2])

        index2.append(item[3])

        index2.append(item[4])

        index3.append(item[5])

        index3.append(item[6])

        index4.append(item[7])

        index4.append(item[8])

        # index5.append(item[9])

        # index5.append(item[10])

# print(index1)

# print(index2)

# print(index3)

# print(index4)

# 数据归一化

index1 = np.array(index1)

index2 = np.array(index2)

index3 = np.array(index3)

index4 = np.array(index4)

# index5 = np.array(index5)

indexNew1 = [(i-min(index1))/(max(index1)-min(index1)) for i in index1]

indexNew2 = [(i-min(index2))/(max(index2)-min(index2)) for i in index2]

indexNew4 = [(i-min(index4))/(max(index4)-min(index4)) for i in index4]

# 距离 发球失误是负指标，所以需要取max-x

# indexNew3 = [(max(index3)-i)/(max(index3)-min(index3)) for i in index3]

# indexNew3 = [0.