Code

109611099_hw4_resubmit.py - C:/prog/ML/hw4/109611099_hw4_resubmit.py (3.8.3)

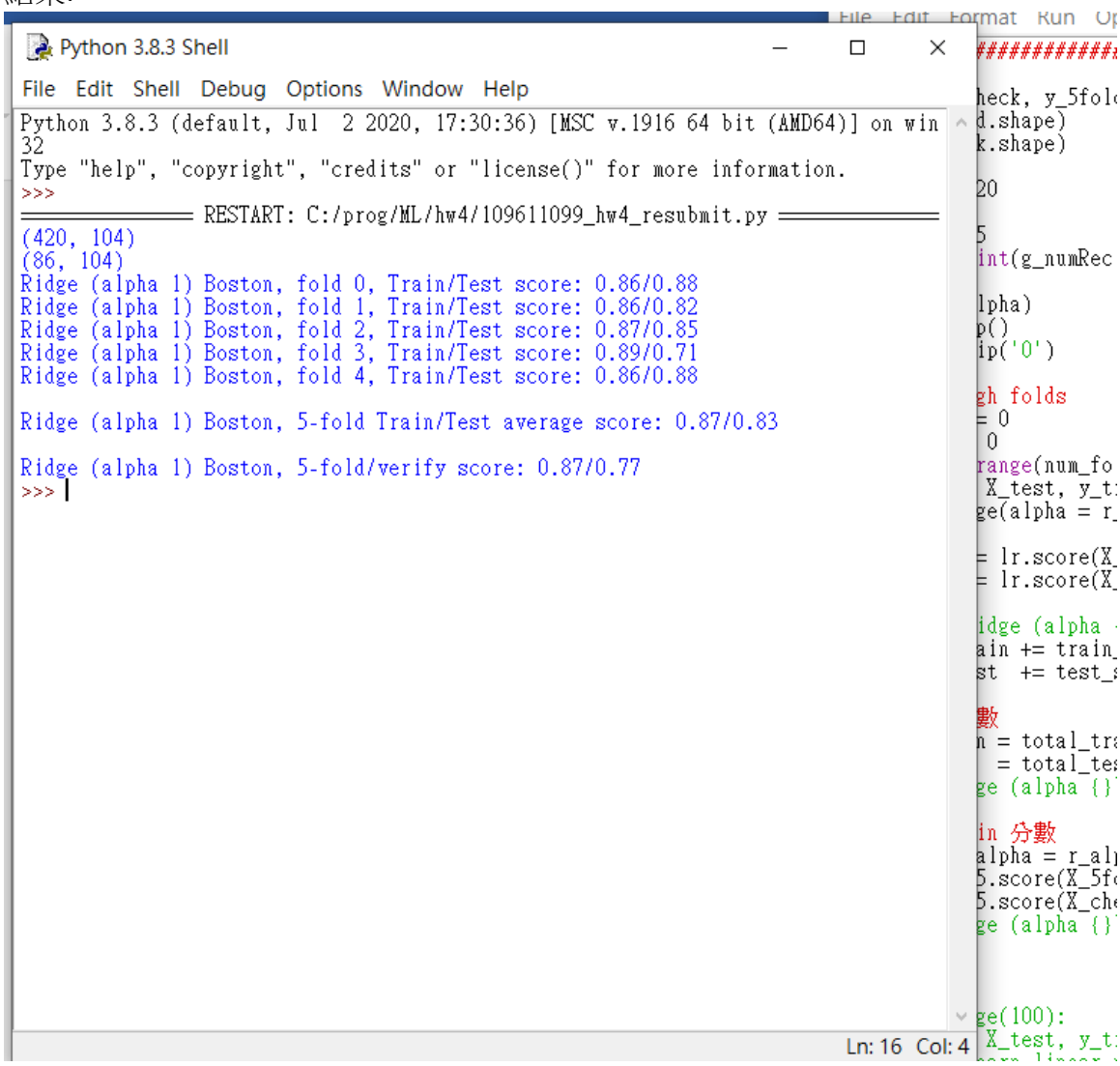
File Edit Format Run Options Window Help

```
import numpy as np
import mglearn
import pandas as pd
import sklearn
import matplotlib.pyplot as plt
from sklearn.linear_model import Ridge

import random
from sklearn.model_selection import train_test_split

import numpy as np
def Read(indata):
    #init
    recArr = []
    clsArr = []

    #read file
    f = open(indata, 'r')
    row = 0
    while (row < 506):
        s = f.readline()
        data1 = s.strip() # remove leading and ending blanks
        if (len(data1) <= 0):
            break

        # since we use append, value must be created in the loop
        val=[0.]
        size=104
        value=[val]*104
        target=[0.]

        data1 = data1.replace('[', '') # remove [
        data1 = data1.replace(']', '') # remove ]
        strs105 = data1.split()  # array of 2 str

        # convert to real
        for i in range(104):#0-103 共104各
            value[i] = eval(strs105[i])
        target[0]=eval(strs105[104])

        #print("row = {}".format(row) + ", {}\n".format(value), end='')

        recArr.append(value) ; # add 1 record at ending
        clsArr.append(target)

        row = row+1 # total read counter

    # close input file
    f.close()

    npXY = np.array(recArr)
    npXY=npXY.reshape(506,104)
    npC  = np.array(clsArr)
    npC=npC.reshape(506,)
    return npXY, npC
############################################################
```

```
#########################################################
def TrainTestSplit_Fold(X, y, fold, test_size):
    # safety check
    if (fold < 0):
        fold = 0

    # 輸入
    numValue = X.size
    rows = len(X)
    cols = int(numValue/rows)

    # safety check
    rmns = numValue % rows
    if (rmns != 0):
        print("ERROR - missing data in X")

    t0 = fold * test_size
    t1 = t0 + test_size
    # safety check
    if (t1 > rows):
        print("ERROR - out of bound")
        t1 = rows


    fea_test = X[t0:t1,:]
    tar_test = y[t0:t1]


    dr = [t0+x for x in range(test_size)]


    fea_train = np.delete(X, dr, 0)
    tar_train = np.delete(y, dr, 0)
    return fea_train, fea_test, tar_train, tar_test
# end function

#########################################################
#讀檔
f= "C:\\prog\\ML\\hw4\\hw4_boston.csv"
x,y= Read(f)
#########################################################
#main
X_5fold, X_check, y_5fold, y_check = train_test_split(x, y, test_size = 86, random_state = 0)
print(X_5fold.shape)
print(X_check.shape)
#分割
g_numRec = 420

num_folds = 5
test_size = int(g_numRec * (1.0/num_folds))
r_alpha = 1
rs = str(r_alpha)
rs = rs.strip()
rs = rs.rstrip('0')

# loop through folds
total_train = 0
```

```
# loop through folds
total_train = 0
total_test = 0
for fold in range(num_folds):
    X_train, X_test, y_train, y_test = TrainTestSplit_Fold(X_5fold, y_5fold, fold, test_size)
    lr = Ridge(alpha = r_alpha).fit(X_train, y_train) # y_train is 1-dim array

    train_s = lr.score(X_train, y_train)
    test_s  = lr.score(X_test,  y_test)

    print("Ridge (alpha {}) Boston, fold {}, Train/Test score: {:.2f}/{:.2f}".format(rs, fold, train_s, test_s))
    total_train += train_s
    total_test  += test_s

# 計算平均分數
average_train = total_train/num_folds
average_test  = total_test/num_folds
print("\nRidge (alpha {}) Boston, 5-fold Train/Test average score: {:.2f}/{:.2f}".format( rs, average_train, average_test))

# 5-fold train 分數
lr5 = Ridge(alpha = r_alpha).fit(X_5fold, y_5fold)
fold5_s = lr5.score(X_5fold,  y_5fold)
check_s = lr5.score(X_check,  y_check)
print("\nRidge (alpha {}) Boston, 5-fold/verify score: {:.2f}/{:.2f}".format(rs, fold5_s, check_s))
```

結果:


Python 3.8.3 Shell
File Edit Shell Debug Options Window Help
Python 3.8.3 (default, Jul 2 2020, 17:30:36) [MSC v.1916 64 bit (AMD64)] on win
32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: C:/prog/ML/hw4/109611099_hw4_resubmit.py ==========
(420, 104)
(86, 104)
Ridge (alpha 1) Boston, fold 0, Train/Test score: 0.86/0.88
Ridge (alpha 1) Boston, fold 1, Train/Test score: 0.86/0.82
Ridge (alpha 1) Boston, fold 2, Train/Test score: 0.87/0.85
Ridge (alpha 1) Boston, fold 3, Train/Test score: 0.89/0.71
Ridge (alpha 1) Boston, fold 4, Train/Test score: 0.86/0.88
Ridge (alpha 1) Boston, 5-fold Train/Test average score: 0.87/0.83
Ridge (alpha 1) Boston, 5-fold/verify score: 0.87/0.77
>>>
Ln: 16 Col: 4