

ML HW2

Q3.1 :

這題是要自行輸入k-NN classifier 的k值，先讀取iris.data的資料，將此150筆資料前面四個 attributes轉成float型態並append到trainingDtaSet，最後label的部分用trainingLabelSet 陣列來存，前50筆資料將陣列值設為0，第51~100筆資料陣列值設為1，第100~150筆資料 陣列值設為2，方便後面做運算。

我define了三個function來進行運算，分別為dis、takeSecond、classifier1。

dis (ex1, ex2) :

這是用來算距離的function，將兩個data set傳入，實作距離公式後return 距離。

takeSecond (elem) :

這在後面用到sort()時會使用到，傳入一個list，回傳list第二個值。

classifier1 (k, ex) :

這是主要實作K-NN的function，傳入k值及test data，用一個for-loop跑150次，算出test data與每個training data的距離，並append [training data的label, 距離]到ansset，由於我們 是要找其中k個距離最近的點，於是用ansset.sort(key=takeSecond)讓距離由小排到大，接 著用一個for-loop跑k次，計算前k小的距離對應的label個數（ans[label]=label個數），最後 回傳ans。

最終由user輸入k值，及自行創造的data，呼叫classifier1 (k, testdata)後利用for-loop找出 ans最大值的陣列id，id=0表示第一種，id=1表示第二種，id=2表示第三種，最後輸出其 label，如右圖。

```
: k=input('k:')
k = int(k)
data1 = input('test data1:')
data2 = input('test data2:')
data3 = input('test data3:')
data4 = input('test data4:')
test = [float(data1), float(data2), float(data3), float(data4)]
a = classifier1(k, test)
anskey = 0
f=0
for i in range(3):
    if a[i]>anskey:
        anskey = a[i]
        f = i
if f==0:
    print("Label : Iris-setosa")
elif f==1:
    print("Label : Iris-versicolor")
else:
    print("Label : Iris-virginica")

k:10
test data1:2
test data2:4
test data3:5
test data4:3
Label : Iris-virginica
```

Q3.2 :

這題讀取資料的方式與第一題相同，不同的地方在於將三種型態的前30筆資料append到training的dataSet，後20筆資料append到test的dataSet。

這題我define了兩個function來進行運算，分別為dis、classifier1。

dist (ex1, ex2)：與第一題同。

classifier1(ex)：由於這題是指定使用1-NN的classifier，利用for-loop讓test Data去和90筆training data進行比對，return距離test data最近的training data的label。

最終用for-loop將60筆test data傳入classifier1並利用回傳的label去和正確的label進行比對，相同則correct數往上加，不同則error數往上加，最後印出正確及錯誤的數量，還有正確率，如右圖。

```
In [45]: correct = 0
error = 0
for t in testSet:
    origin = t[4]
    get = classifier1(t)
    if (origin==get):
        correct+=1
    else:
        error+=1
correct_rate = correct/60
print("correct number: %d\n" %correct)
print("incorrect number: %d\n" %error)
print("accuracy: %f\n" %correct_rate)
```

correct number: 58

incorrect number: 2

accuracy: 0.966667