

# **Machine Learning**

## **Programming Assignment 2**

*Name :* 姜博瀚

*ID :* 410421214

GITHUB: [https://github.com/chiangpohan/ML2018\\_HW2\\_410421214](https://github.com/chiangpohan/ML2018_HW2_410421214)

## 一、程式

```
from sklearn.datasets import fetch_mldata
mnist = fetch_mldata("MNIST original")

#download dataset

X, Y = mnist.data, mnist.target

X, Y = shuffle(X, Y)
```

下載 Mnist 的手寫辨識資料後，將其資料 shuffle，避免每次訓練出的結果都相同。

```
pca = PCA(svd_solver="arpack", n_components=69)
X = pca.fit_transform(X)
```

將資料降成 69 維度

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=1/10, random_state=None)

X_train = X_train.astype("float") / 255 * 2 - 1
X_test = X_test.astype("float") / 255 * 2 - 1

#set value from 1 to -1
```

本次訓練將 10% 的資料作為 testing data  
並且將資料向量化至 1~-1

```

clf = NuSVC()
clf.fit(X_train, Y_train)
TEST=clf.predict(X_test)
TRAIN=clf.predict(X_train)

#training dataset

count=0
for i in range(Y_train.shape[0]):
    if TRAIN[i]==Y_train[i]:
        count=count+1

print(count/Y_train.shape[0])

#training data Accuracy

count=0
for i in range(Y_test.shape[0]):
    if TEST[i]==Y_test[i]:
        count=count+1

print(count/Y_test.shape[0])

#testing data Accuracy

```

---

使用 NuSVC 下去做訓練，並印出訓練資料和測試資料的準確率

程式執行結果

```

(63000, 69)
(63000,)
(7000, 69)
(7000,)
0.9454126984126984
0.9428571428571428

```

0.945412 為 訓練資料準確率

0.942857 為 測試資料準確率

## 二、遇到問題

- (1)因為不熟 NuSVC ，在查資料上花了比較多時間。
- (2)由於一開始將 compoment 設為 95，但由於本身 CPU 不是很好，訓練了一個多小時小黑窗仍無任何動靜，且 CPU 佔了太多的使用量。於是重新調整為 69，讓速度快一些，但結果可能就沒那麼好。
- (3)有嘗試參考類似 SGD 畫圖，但我的程式似乎因沒有 Epoch 沒有辦法呈現。

### 三、心得

這次是第一次使用 scikit learn 的 NuSVC 。由於之前較不熟悉，只有聽過類似的介紹。於是乎打開了 sklearn 的網站上去找他的詳細資訊，包含他的預設和其他函式要如何呼叫等等。

在上學期在上顏老師的人工智慧時，老師有在課堂上提過 Keras 的手寫辨識大概的內容，以及如何去寫。但是那堂課並未實作那部分，所以這次打程式時花了很多時間在找資料和詢問同學一些細節怎麼去設比較好。

#### 四、結論

這次的成果有約 94% 的準確率，對第一次實作的我看來算是不錯的，畢竟之前沒有任何的相關經驗。但是如果這是要拿出去市面的程式來看，94% 應該是相當不足的，可能需要再調整 component (亦或是更換 CPU 去讓他跑更高的維度測試)，或是作其他修正準確率才會有更大的提升。