INTRODUCTION TO

NEURAL NETWORKS

類神經網路簡介

Final Project 1

~ Asthma Patient Identification ~

|  |  |
| --- | --- |
| 組員 : | N16041422 陳友政 |
|  | N16041448 黃子軒 |
|  | N16054629 張元睿 |
|  | N16041749 侯均憲 |

授課老師 : 王振興 教授

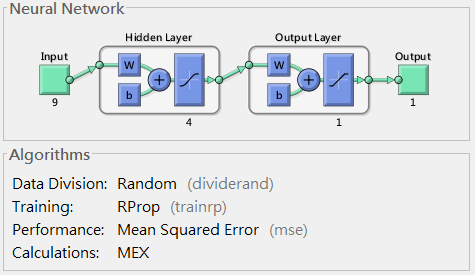
**分類器1：Back Propagation Neural Network**

**流程與方法:**

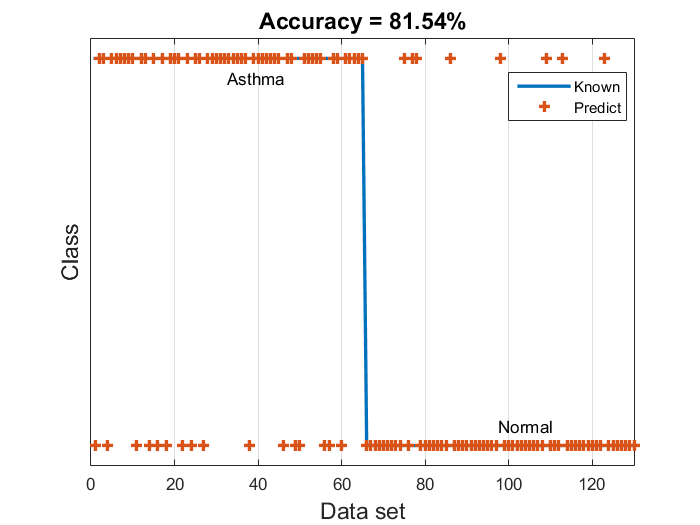
1. 以Sequential feature selection篩選參數，先用k-fold ( k = 25)區分實驗組，再以sequentialfs函數輸出篩選結果，共作50次，統計所有結果並輸出各參數出現機率，選出高於機率高於10%的參數。
2. 區分參數組別，將資料分為(1)訓練組( n = 100)及(2)測試組( n = 30)，先隨機選出訓練組的Index，再以setdiff函數算出測試組的Index。設定Step 1.選出的參數後，用k-fold ( k = 10)選出類神經中每次測試的組別，執行下一步驟。
3. 設定訓練各項參數，本報告設定參數如下：

|  |  |  |
| --- | --- | --- |
| **Hidden layer** | 4 neuron | tansig |
| **Output layer** | 1 neuron | satlins |
| **Algorithm** | trainrp | |
| **Learning rate** | 0.000001 | |
| **DivideFcn** | divideind | |

1. 反覆執行Step 2.至Step.3求出最佳辨識率。
2. 以測試組代入訓練好的網絡，並輸出辨識率與辨識結果。



**預測結果：**



**問題討論:**

1. 代入自行選擇的參數進行訓練，並訓練效果並不如預期，可能原因為資料的差異性並不顯著。
2. 資料的品質不佳，這會導致訓練效率不彰。
3. 可能需辨換測量的方式，或是需挑選顯著性較高的參數，辨識結果才有機會到90%。

**分類器2：Support Vector Machine**

**流程與方法:**

1. 讀取檔案
2. Feature Selection
3. 篩選出最適合的參數
4. 複製Data Set並打散複製後Data Set順序
5. 將參數正規化
6. 將Data Set分為train set以及valid set
7. Train SVM並計算辨識率

**預測結果：**

Train 5 次紀錄train set辨識率以及valid辨識率如下表

|  |  |  |
| --- | --- | --- |
| K-Fold Loss | Train Rate | Valid Rate |
| 0.125 | 93.99% | 93.26% |
| 0.111 | 94.23% | 92.31% |
| 0.120 | 93.51% | 91.35% |
| 0.099 | 94.47% | 83.65% |
| 0.130 | 9.40% | 85.58% |

**問題討論:**

1. Feature selection的函式和SVM辨識結果不同，所選出來的feature不一定是最好的
2. 複製Data Set並打散複製後Data Set順序，是為了讓data set隨機餵給SVM訓練時不受順序影響，但重複性的train data會兆呈over fitting
3. Matlab KFold 都能在0.15以下，反映出data set以及所選feature都能一定程度辨識出氣喘與否