

Assignment #3 –Adaptive Boosting

資工三 409410114 周述君

1. List all parameters that should be set before running Adaboost. Explain the meanings of those parameters.
 - Train: 用來訓練 Adaboost 模型的樣本
 - train_label: training dataset 中樣本的標籤
 - test: 評估模型效能的樣本
 - test_label: test dataset 中樣本的標籤
 - step: 決定每個 iteration 所增加的訓練樣本數量或增加週期循環次數
 - cycles: 決定 Adaboosting 演算法會應用在 training dataset 的次數
 - maxCycles: 迴圈執行次數的上限
 - errorTrain: Adaboosting 模型錯誤分類的 training dataset 百分比
 - errorTest: Adaboosting 模型錯誤分類的 test dataset 百分比

```
11 function runAdaBoosting(train,train_label,test,test_label)
12     disp('run adaboost with cycles=100');
13     step=100;
14     cycles=100;

40     step=20;
41     maxCycles=100;
42     for cycles=step:step:maxCycles
43         disp(cycles);
44         boost=adaBoost(train,train_label,cycles);
45         [errorTrain(cycles/step),errorTest(cycles/step)]=getError(boost,train,train_label,test,test_label);
46         clear boost;
47     end
```

2. How each weak learner is decided and trained in each iteration? What is the learning algorithm A? Does it use bootstrapped dataset? If not, how Dt is obtained for each iteration?

Adaboost 的每個 iteration 都會選擇並訓練一個弱學習器把訓練集的誤差最小化。Learning algorithm A 是用決策樹實作(root 和 leaves 直接連接)。並非使用 bootstrapped dataset，Dt 的分佈會依每此迭代時前一個弱學習器的性能進行更新，若在前一次迭代被錯誤分類就會增加權重，以提高下次被弱學習器選擇的機會，因此模型會變成弱學習器的加權組合。

```
4 for tt=1:(16*256-1)
5     error(tt)=distribution*abs(label-(train(:,floor(tt/16)+1)>=16*(mod(tt,16)+1)));
6 end
```

3. List the first three weak learners when the learning iteration stops. Explain these decision stumps by their three parameters i , θ and s .

| | 1 | 2 | 3 |
|---|--------|----------|---------|
| 1 | 0.3774 | 11.0000 | 80.0000 |
| 2 | 0.4807 | 170.0000 | 80.0000 |
| 3 | 2.0170 | 58.0000 | 16.0000 |

```
prediction=prediction-log(boost(h,1))*(train(:,boost(h,2))>=boost(h,3));
```

看混和權重的第 i 個維度是否大於閾值，再判斷預測是否大於-
sum(log(weighted sum))

a popular choice: decision stump

- $$h_{s,i,\theta}(\mathbf{x}) = s \cdot \text{sign}(x_i - \theta)$$
 - positive and negative rays on some feature: three parameters (feature i , threshold θ , direction s)
 - physical meaning: vertical/horizontal lines in 2D
 - efficient to optimize: $O(d \cdot N \log N)$ time

$$h_{s,i,\theta}^1(x) = 0.3774 * \text{sign}(x_{11} - 80)$$

$$h_{s,i,\theta}^2(x) = 0.4807 * \text{sign}(x_{170} - 80)$$

$$h_{s,i,\theta}^3(x) = 2.0170 * \text{sign}(x_{58} - 16)$$

i : 決定特徵，表示在決策過程中考慮哪個功能

θ : 分割數值的閾值，決定點的分類

s : 決定分向，看大於/小於閾值決定該被分類在正/負

4. Following 3), list the blending weights of these three decision stumps. Explain how their blending weights are decided and what are their actual values in the program

$$\text{blending weight} = \ln\left(\frac{\text{error}}{1 - \text{error}}\right) = \ln\left(\frac{1 - \epsilon_t}{\epsilon_t}\right)$$

- prediction

| | 1 |
|---|---------|
| 1 | 0.3886 |
| 2 | 3.5991 |
| 3 | -1.0113 |
| 4 | -0.7509 |
| 5 | 1.7248 |
| 6 | 2.5088 |
| 7 | -0.3068 |
| 8 | 1.4511 |
| 9 | 1.2253 |

- temp

| | |
|---|--------|
| | 1 |
| 1 | 1.1206 |

判斷 prediction 結果有沒有大於 temp

Result

