

機器學習< Assignment #3 - Adaptive Boosting>

姓名：翁佳煌

學號：409430030

問題 1. List all parameters that should be set before running Adaboost.
Explain the meanings of those parameters.

train: 訓練集，即一個矩陣，每行代表一個樣本，每列代表一個特徵。

train_label: 訓練集標籤，即一個向量，包含每個樣本對應的標籤。

cycles: 迭代次數，即要訓練幾個弱分類器。

```
function boosted=adaBoost(train,train_label,cycles)
    disp('running adaBoost algorithm');
    d=size(train);
    distribution=ones(1,d(1))/d(1); %每一個訓練樣本最開始時都被賦予相同的權值：1/N。
    error=zeros(1,cycles);
    beta=zeros(1,cycles);
    label=(train_label(:)>=5);% contain the correct label per vector
    for j=1:cycles
        if(mod(j,10)==0)
            disp([j,cycles]);
        end
        [i,t]=weakLearner(distribution,train,label);
        error(j)=distribution*abs(label-(train(:,i)>=t)); % 是在第j次迭代中弱分類器的錯誤率
        beta(j)=error(j)/(1-error(j)); %beta(j)表示第j次迭代中的弱分類器的權重 和error成反比 代表錯誤率越小權重越大
        boosted(j,:)=[beta(j),i,t];
    end
end
```

問題 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 D_t is obtained for each iteration?

在每一次的迭代中，弱分類器是通過執行 weakLearner 函數來獲取一個弱分類器，這個弱分類器是基於當前的權重分佈（即每個樣本的權重）來訓練的。

weakLearner 函數實現了一個二元分類器，使用了所有訓練樣本的權重分佈 distribution。在每一次迭代中，算法都會根據當前樣本權重分佈，從給定的弱分類器集合中選擇一個表現最好的弱分類器。

```
function [i,t] = weakLearner(distribution,train,label)
    %disp('run weakLearner');
    for tt=1:(16*256-1)
        error(tt)=distribution*abs(label-(train(:,floor(tt/16)+1)>=16*(mod(tt,16)+1)));
    end
    [val,tt]=max(abs(error-0.5));

    i=floor(tt/16)+1;
    t=16*(mod(tt,16)+1);
```

在這裡並沒有使用 bootstrapped dataset，而是使用了在訓練數據上的初始分佈 distribution 來調整每個樣本的權重，以便在下一次迭代中更關注被錯誤分類的樣本。每次迭代後，distribution 會被更新以反映每個樣本的新權重。

```
[i,t]=weakLearner(distribution,train,label);
error(j)=distribution*abs(label-(train(:,i)>=t)); % 是在第j次迭代中弱分類器的錯誤率
beta(j)=error(j)/(1-error(j)); %beta(j)表示第j次迭代中的弱分類器的權重 和error成反比 代表錯誤率越小權重越大
boosted(j,:)=[beta(j),i,t];

distribution=distribution.* exp(log(beta(j))*(1-abs(label-(train(:,i)>=t))))');
% 計算當前迭代得到的弱分類器的加權預測結果，用於更新每個樣本的權重distribution。
% 對於每個樣本，如果它被弱分類器正確分類，那麼它的權重就會下降；如果它被錯誤分類，那麼它的權重就會上升。
% 這樣，被分錯的樣本在下一次迭代中就會更有可能被選中，從而讓後面的弱分類器更關注這些被分錯的樣本。

distribution=distribution/sum(distribution); %Normalization每個資料點的權重以保證它們總和為1

end
```

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

我在程式碼最後面加了下面的 code 去做觀察，印出對應的 i 、 θ 和 s 。

```
for k = 1:3
    disp(['Iteration ', num2str(k) ': i=' num2str(boosted(k,2)) ', t=' num2str(boosted(k,3)) ', s=' num2str(sign(boosted(k,1)))]);
end
```

```
Iteration 1: i=170, t=160, s=1
Iteration 2: i=26, t=192, s=1
Iteration 3: i=74, t=144, s=1
```

i - 用於拆分數據的特徵的索引。

θ - 用於分割數據的閾值。

s - 分類方向，正向 ($s = 1$) 或負向 ($s = -1$)。

例如，第一個弱分類器的參數是 $i=170$ ， $\theta=160$ ， $s=1$ 。這個弱分類器將第 $i=170$ 個特徵作為門檻值 ($t=160$)，將輸入向量分為兩個類別。如果輸入向量的第 i 個特徵大於等於閾值 t ，那麼它被分類為正例 (+1)，否則被分類為負例 (-1)。

問題 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?

每個弱分類器都會得到一個權重，這個權重反映了它對最終分類器的貢獻程度。權重越大，對分類結果的影響就越大。在 AdaBoost 算法中，權重是根據每個弱分類器的錯誤率計算得出的。錯誤率越小，權重越大。

對於每個弱分類器，在計算出它的分類錯誤率後，會得到一個權重係數 $\beta(j)$ 。然後，每個樣本的權重會按照以下公式進行更新：

$\text{distribution} = \text{distribution} * \exp(\log(\beta(j)) * (1 - \text{abs}(\text{label} - (\text{train}(:, i) \geq t))))$;

這裡的 distribution 表示每個樣本的權重， train 表示訓練樣本， label 表示每個樣本的真實標籤， i 和 t 表示弱分類器的參數。更新後，每個樣本的權重會重新歸一化，以保證它們總和為 1。

```
for k = 1:3
    disp(['Blending weight of weak learner ' num2str(k) ': ' num2str(beta(k))]);
end
```

```
Blending weight of weak learner 1: 0.33333
Blending weight of weak learner 2: 0.2931
Blending weight of weak learner 3: 2.7384
```