CS 559: Homework Set 4

Alana Laryssa S A Santos

12/3/2015

1 First problem

$$Y = \begin{bmatrix} 1 & -2 & 1 \\ 1 & -5 & -4 \\ 1 & -3 & 1 \\ 1 & 0 & -3 \\ 1 & -8 & -1 \\ -1 & -2 & -5 \\ -1 & -1 & 0 \\ -1 & -5 & 1 \\ -1 & 1 & 3 \\ -1 & -6 & -1 \end{bmatrix}$$

The final weight vector is $a = \begin{bmatrix} -0.1010 & -0.1805 & -0.270 \end{bmatrix}^T$. Only two points were misclassified: $\begin{bmatrix} 0 & -3 \end{bmatrix}$ and $\begin{bmatrix} -1 & -3 \end{bmatrix}$. The rest of the points were classified correctly.

1.1 MatLab code

```
D1 = [-2 1; -5 -4; -3 1; 0 -3; -8 -1];
D2 = [2 5; 1 0; 5 -1; -1 -3; 6 1];

b = 1;

ld1 = size(D1,1);
ld2 = size(D2,1);

bb = repmat(b,ld1+ld2,1);

Y = [ones(ld1,1) D1;
     -ones(ld2,1) -D2];

a = inv(Y'*Y)*Y'*bb;

% figure, scatter(D1(:,1), D1(:,2), 'b', '*')
% hold on, scatter(D2(:,1), D2(:,2), 'm', '*')
% grid on; axis equal;

d = [D1;D2];
```

```
tmp = sum([a(2)*d(:,1) a(3)*d(:,2) repmat(a(1),length(d),1)],2);
idx = double(tmp > 0);
idx(idx==0) = -1;

correct = sum(idx==Y(:,1));
disp(['correct: ' num2str(correct) '/' num2str(size(Y,1))])
```

2 Second problem

The margin is given by $m = 2/||w||$ and we want the distance to the closest samples to be large. In order to do that, we have to minimize $||w||$. Not mathematically, no. But we can minimize $\frac{1}{2}||w||^2$ instead.

3 Third problem

```
$ ./train -v 5 -c 1 ../pima2.data

optimization finished, #iter = 1
Objective value = -615.000000
nSV = 615

optimization finished, #iter = 1
Objective value = -614.000000
nSV = 614

optimization finished, #iter = 1
Objective value = -615.000000
nSV = 615

optimization finished, #iter = 1
Objective value = -614.000000
nSV = 614

optimization finished, #iter = 1
Objective value = -614.000000
nSV = 614
Cross Validation Accuracy = 34.8958%
```

Where ../pima2.data was generated with the MatLab code below:

```
data = dlmread('pima-indians-diabetes.data');
data = [data(:,9) data(:,2:4)];
```

```
fid = fopen('pima2.data');
fprintf(fid, '%d %d %d %d\n', data');
fclose(fid);
```

4 Fourth problem

Accuracy in boosting rounds:

- 1st: 75%
- 2nd: 65.5%
- 3rd: 65.36%
- Training accuracy: 67.06%

4.1 MatLab code

```
clear all;
data = dlmread('pima-indians-diabetes.data');

x = data(:,2:4);
y = data(:,9); y(y==0) = -1;

len = length(x);

% adaboost
w = repmat(1/len,len,1);
cnt = 1;

for i = ['x' 'y' 'z']
    [eps,h,thres] = weakLearner(i,x,y,50,w);

    alpha = .5*log((1-eps)/eps);

    w = w.*exp(-alpha.*y.*h);
    w = w/sum(w);   % w/z

    finalh(cnt) = thres;
    finalalpha(cnt) = alpha;
    cnt = cnt + 1;

    1-eps
end
```

```
%% to test:
%% H(x) =
%       tmp = x > repmat(finalh,len,1);
%       tmp = double(tmp);
%       tmp(tmp==0) = -1;
%       h = sign(sum(tmp.*alpha,2));
%       acc = sum(h==y)/len


function [minerr,minh,minthres] = weakLearner(axiss,data,lbl,steps,weight)

    if strcmp(axiss,'x')
        mode = 1;
    elseif strcmp(axiss,'y')
        mode = 2;
    elseif strcmp(axiss, 'z')
        mode = 3;
    else
        return;
    end

    maxy = max(data(:,mode));
    miny = min(data(:,mode));

    thres = miny : (maxy-miny)/steps : maxy;

    for i = 1:length(thres)
        idx = data(:,mode) >= thres(i);
        idx = double(idx);
        idx(idx==0) = -1;
%         err = exp(-idx.*lbl);
        err = sum(idx~=lbl)/length(data);
        tempeps(1,:) = sum(weight.*err);
        temph(1,:) = idx';

        idx = data(:,mode) < thres(i);
        idx = double(idx);
        idx(idx==0) = -1;
%         err = exp(-idx.*lbl);
        err = sum(idx~=lbl)/length(data);
        tempeps(2,:) = sum(weight.*err);
        temph(2,:) = idx';

        [~,minidx] = min(tempeps);
        eps(i,:) = tempeps(minidx,:);
        h(i,:) = temph(minidx,:);
```

```
    end

    [minerr, minidx2] = min(eps);
    minh = h(minidx2,:)';
    minthres = thres(minidx2);

%       figure, scatter(data(lbl<0,1), data(lbl<0,2), '*', 'r');
%       hold on; scatter(data(lbl>0,1), data(lbl>0,2), '*', 'b');
%       grid on; axis equal;
%       if mode == 1
%           yy = min(data(:,1))-1:max(data(:,1))+1;
%           xx = repmat(thres(minidx2),1,size(yy,2));
%       else
%           xx = min(data(:,1))-1:max(data(:,1))+1;
%           yy = repmat(thres(minidx2),1,size(xx,2));
%       end
%       hold on, plot(xx,yy);

end
```