COMS 4721 Machine Learning for Data Science Name: Sung-Yen Liu UNI: sl3763

HW#5

1.

Code:

```
% hw5.1

load('./hw5mat/CFB2014.mat');

M = zeros(759,759);

for i = 1:size(scores,1)
        pj1 = scores(i,2)/(scores(i,2)+scores(i,4));
        pj2 = scores(i,4)/(scores(i,2)+scores(i,4));

        M(scores(i,1), scores(i,1)) = M(scores(i,1), scores(i,1)) + cpr(scores(i,2), scores(i,4)) + pj1;
        M(scores(i,3), scores(i,3)) = M(scores(i,3), scores(i,3)) + cpr(scores(i,4), scores(i,2)) + pj2;
        M(scores(i,1), scores(i,3)) = M(scores(i,1), scores(i,3)) + cpr(scores(i,4), scores(i,2)) + pj2;
        M(scores(i,3), scores(i,1)) = M(scores(i,3), scores(i,1)) + cpr(scores(i,2), scores(i,4)) + pj1;

end

normM = M ./ repmat(sum(M,2),[1 759]);
% sum(normM(1,:))
%%%

w0 = repmat(1/759,[759 1]);
wrec = zeros(759,4);
value = zeros(20,4);
ranking = zeros(20,4);
rank_name = cell(20,4);
wite = [10, 100, 200, 1000];

for i = 1:4
        wcur = w0;
        for j = 1:wite(i)
                wnext = (normM')*wcur;
                wcur = wnext;
        end
        wrec = wcur;
```

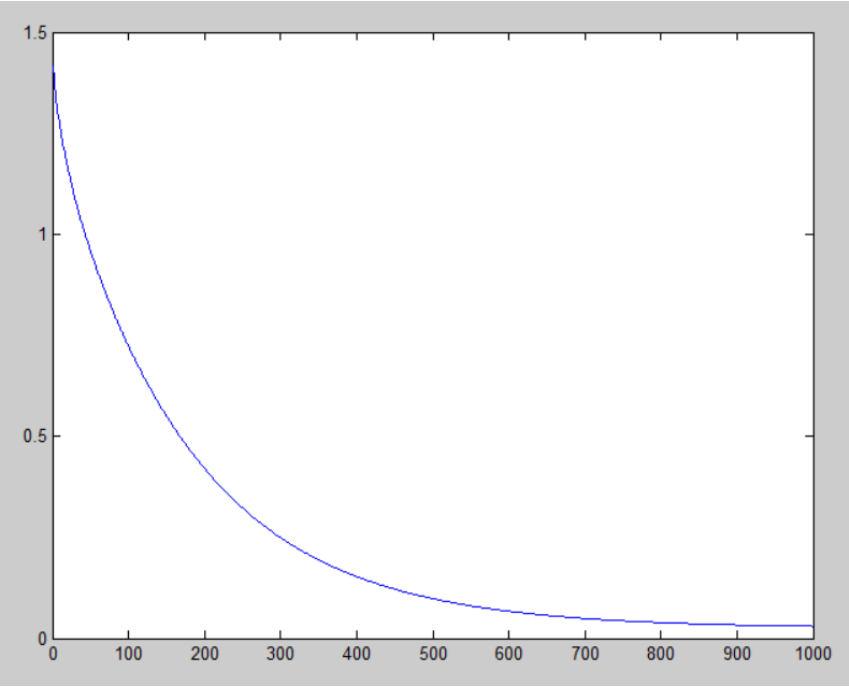
```
        [t1, t2] = sort(wnext,'descend');
        value(:,i)=t1(1:20);
        ranking(:,i)=t2(1:20);
        rank_name(:,i) = legend(ranking(:,i));
end

[egvc, eg] = eig(normM');
error = zeros(1000,1);
%%
egg = zeros(759,1);
for i = 1:759
        egg(i) = eg(i,i);
end
%findidx = find(egg==max(egg));
%winf = egvc(:,findidx);
[rr, ii] = sort(egg,'descend');
winf = egvc(:,ii(2));
%%

winf = winf./sum(winf);
wcur = w0;
for i = 1:1000
        wnext = (normM')*wcur;
        error(i) = sum(abs(wnext-winf));
        wcur = wnext;
end
plot(error)
```

Iteration (Ranking)			
10	100	200	1000
'UW-Whitewater'	'UW-Whitewater'	'OhioState'	'OhioState'
'MountUnion'	'OhioState'	'Oregon'	'Oregon'
'ColoradoSt-Pueblo'	'Oregon'	'Alabama'	'Alabama'
'OhioState'	'Alabama'	'TCU'	'TCU'
'Linfield'	'TCU'	'FloridaSt'	'FloridaSt'
'MinnSt-Mankato'	'MountUnion'	'UW-Whitewater'	'MichiganSt'
'Wartburg'	'FloridaSt'	'MichiganSt'	'Baylor'
'Wesley'	'ColoradoSt-Pueblo'	'Baylor'	'GeorgiaTech'
'SouthernOregon'	'MichiganSt'	'GeorgiaTech'	'UCLA'

'Oregon'	'SouthernOregon'	'UCLA'	'Mississippi'
'Alabama'	'Baylor'	'Mississippi'	'Georgia'
'NorthDakotaSt'	'GeorgiaTech'	'Georgia'	'Arizona'
'TCU'	'Wartburg'	'Arizona'	'ArizonaSt'
'MaryHardin-Baylor'	'UCLA'	'ArizonaSt'	'MississippiSt'
'FloridaSt'	'Mississippi'	'MississippiSt'	'Missouri'
'Hobart'	'CarrollMT'	'Missouri'	'Clemson'
'JohnCarroll'	'Georgia'	'Clemson'	'SouthernCal'
'MarianIN'	'Arizona'	'SouthernCal'	'Wisconsin'
'Widener'	'ArizonaSt'	'Wisconsin'	'Auburn'
'Concord'	'MississippiSt'	'Auburn'	'Utah'

Iteration (Corresponding value)			
10	100	200	1000
0.015323	0.029276	0.036262	0.047325
0.013074	0.0276	0.029769	0.038733
0.01026	0.022671	0.024338	0.031659
0.009007	0.018663	0.023997	0.031354
0.008784	0.018262	0.019162	0.025011
0.008212	0.016874	0.018215	0.021938
0.008018	0.014655	0.016919	0.020849
0.007488	0.012998	0.015983	0.020133
0.007438	0.012995	0.015446	0.018138
0.007152	0.012489	0.013954	0.017478
0.007137	0.012186	0.013474	0.016958
0.00687	0.011841	0.013101	0.016863
0.006451	0.010681	0.012994	0.015376
0.005976	0.010638	0.011875	0.015322
0.00594	0.010367	0.011794	0.014135
0.005901	0.010276	0.011082	0.013683
0.005863	0.010133	0.010519	0.013519
0.005799	0.009914	0.010424	0.013336
0.005786	0.009097	0.010409	0.013303
0.005685	0.009065	0.010258	0.012912

$plot\ of\ \left\| w_t - u_1 / \sum_j u_1(j) \right\|_1$



$\left\| w_{1000} - u_1 / \sum_j u_1(j) \right\|_1 = 0.0299$

2.

Part1

Code:

```
% hw5.2 p1
%tic
load('./hw5mat/faces.mat');

w = rand(1024,25);
h = rand(25,1000);


error = zeros(200,1);

for i = 1:200
        tmp1 = w'*X;
        tmp2 = w'*w*h;
        h = h.*(tmp1./tmp2);

        tmp3 = X*h';
```

```
            tmp4 = w*h*h';
            w = w.*(tmp3./tmp4);
            diff = X-w*h;

            error(i) = norm(diff);
end

im3 = [101, 106, 121];
sl1 = X(:,im3(3));
tt = h(:,im3(3));
maxh = find(tt==max(tt));
sl2 = w(:,maxh);
sl1tmp = reshape(sl1,32,32);
sl2tmp = reshape(sl2,32,32);
subplot(1,2,1), imagesc(sl1tmp), title(['image121']);
subplot(1,2,2), imagesc(sl2tmp), title(['highest weight']);
%toc
```

Plot of objective function

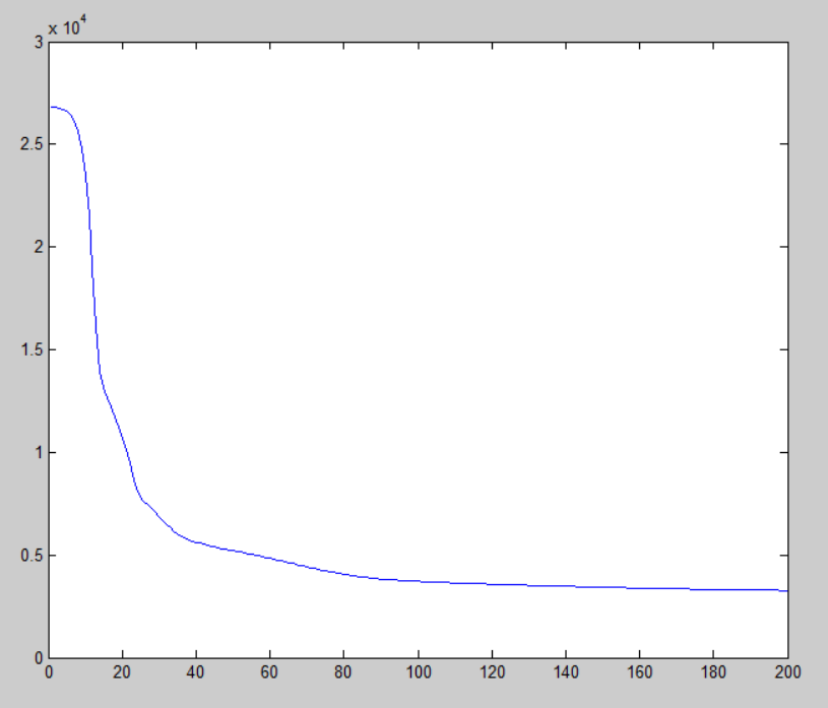


Image 101

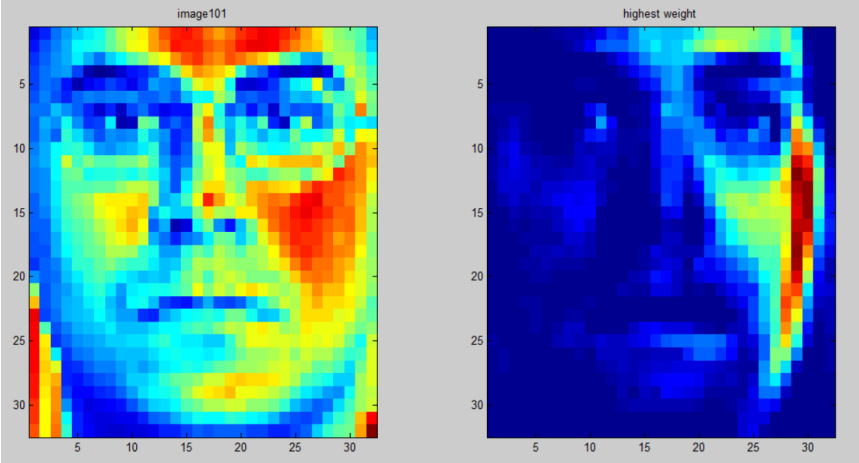


Image 106

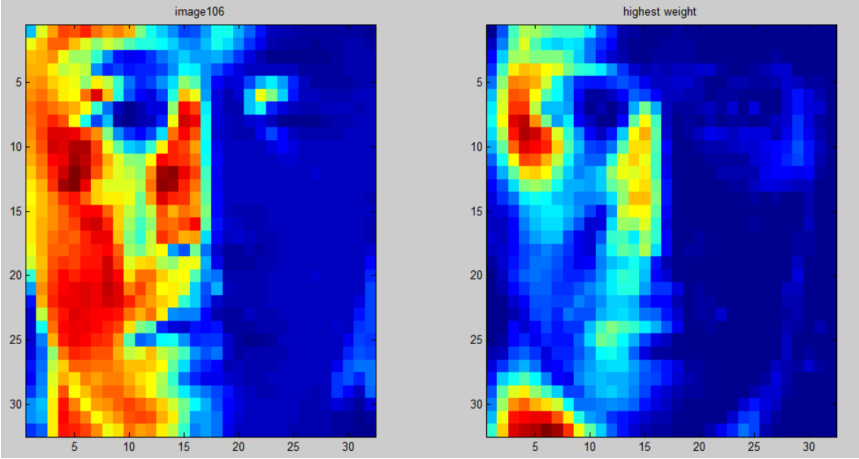
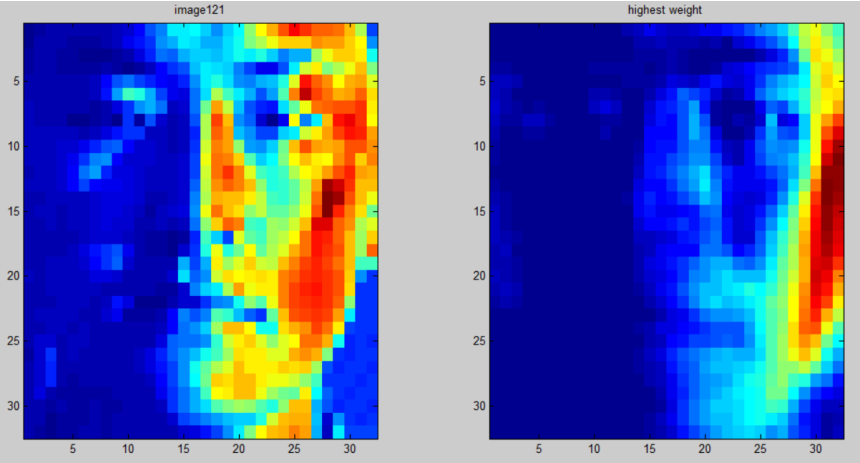


Image121



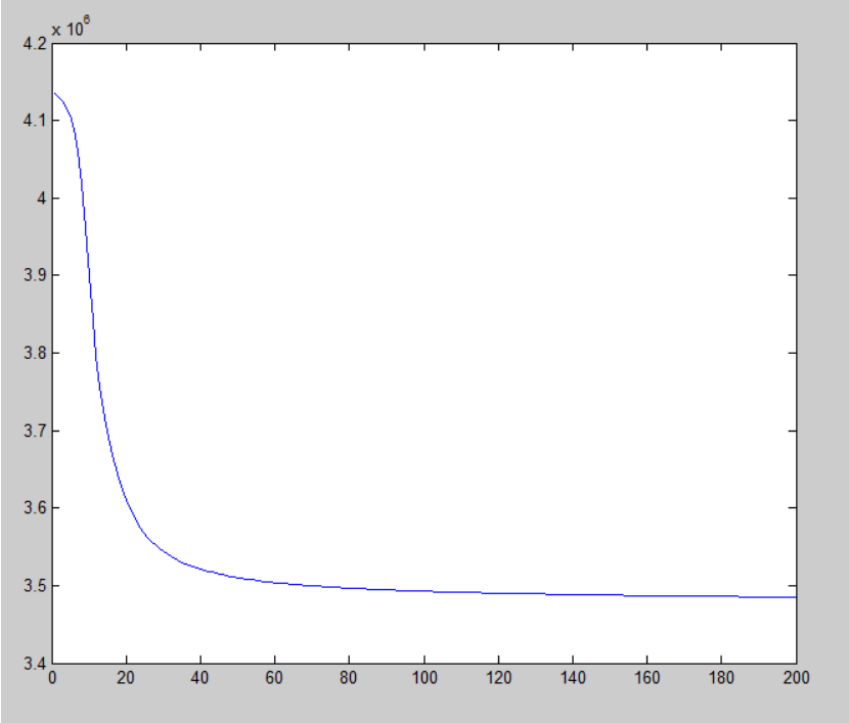
Part2

Code:

```
% hw5.2 p2
tic
load('./hw5mat/nyt_data.mat');

X = zeros(3012,8447);
for i = 1:8447
        tmp1 = Xid{i};
        tmp2 = Xcnt{i};
        X(tmp1,i) = tmp2;
end

w = rand(3012,25);
h = rand(25,8447);
error = zeros(200,1);

eps1 = repmat(eps,[3012 25]);
eps2 = repmat(eps,[25 8447]);
eps3 = repmat(eps,[3012 8447]);

for i = 1:200
        purple = X./ (w*h+eps3);
        pink = (w')./ (repmat(sum(w',2), [1 size(w,1)])+eps1');
        h = h.*(pink*purple);

        purple = X./ (w*h+eps3);
        water = (h')./ (repmat(sum(h'),[size(h',1) 1])+eps2');
        w = w.*(purple*water);

        lg = (log(ones(3012,8447)./ (w*h+eps3)));
        entropy = sum(sum(X.*lg+w*h));
        error(i) = entropy;
end
toc

%%
highestprob = zeros(10,5);
highestvocab = cell(10,5);
```

```
art = [1,2,3,4,5];
normW = w./ repmat(sum(w),[3012 1]);
for i = 1:5
        hi = normW(:,art(i));
        [h1, h2] = sort(hi,'descend');
        highestprob(:,i) = h1(1:10);
        highestvocab(:,i) = nyt_vocab(h2(1:10));
end
```

Plot of objective function



Highest probability and corresponding vocabularies									
1		2		3		4		5	
'thing'	0.016524	'color'	0.00794	'war'	0.018115	'food'	0.017613	'game'	0.041915
'lot'	0.012929	'wear'	0.007067	'military'	0.016473	'restaurant'	0.009581	'team'	0.03948
'feel'	0.011136	'design'	0.006748	'government'	0.013494	'serve'	0.009528	'player'	0.027477
'really'	0.009807	'white'	0.006681	'leader'	0.011752	'water'	0.008931	'season'	0.026413
'little'	0.009783	'wall'	0.006432	'force'	0.011198	'fresh'	0.00854	'play'	0.025078
'tell'	0.009782	'piece'	0.006239	'american'	0.01024	'taste'	0.00835	'coach'	0.014834
'ask'	0.009249	'small'	0.005937	'states'	0.009088	'eat'	0.008249	'league'	0.010553
'big'	0.008964	'paint'	0.005853	'country'	0.009066	'wine'	0.007791	'baseball'	0.010439
'job'	0.00793	'light'	0.005756	'peace'	0.008759	'pound'	0.007643	'ball'	0.009421
'old'	0.007785	'blue'	0.005668	'official'	0.008634	'dry'	0.007476	'football'	0.007806