Assignment 5
MVDA

NA17B111

Manish Patidar

QUETION 1}}

```matlab
load('news_posts.mat');
Z=documents;
[m,n] = size(Z)
%converting a document matrix into full matrix
data=full(double(Z));

 %mean shifting
data=data';
means=mean(data);
for i=1:100
    data(i,:)=data(i,:)-means;
end
%applying svd to get pcs variances
 [U S V]=svds(data,100);
 %variance of first three pcs are
pc1_var=S(1,1)*S(1,1);
pc2_var=S(2,2)*S(2,2);
pc3_var=S(3,3)*S(3,3);
total_var=sum(diag(S).^2);
var_perc=(pc1_var+pc2_var+pc3_var)/total_var;
fprintf('Percentage Variance captured by  first 3 PCs %f\n',var_perc);
fprintf('Variance by 1st PC %f\n',pc1_var);


%-------------------------------------------------
%PART B
for  i=1:100
    [F,adj_var,cum_var] = sparsePCA(data, i, 1);
    if adj_var>pc1_var*(.75)
```

```matlab
        p=i;
        load_vec=F;
        break
    end
end

load_vec;
ind_nonzero=find(load_vec);
%word on that indices
fprintf('No. of Non-zero elements in Sparse PCs for 0.75 of Variance in 1st PC is
%f\n',p);
fprintf('List of words in 1st Sparse PC are \n')
words_list=wordlist(ind_nonzero)




% %-----------------------------------------
%PART C
for  j=1:100
    [F1,adj_var1,cum_var1] = sparsePCA(data,j, 2);
    if sum(cum_var1(2))>=(pc1_var+pc2_var)*(.75)
        k=j;
        load_vec2=F1;
        break
    end
end

% load_vec2
 ind_nonzero1=find(load_vec2(:,2));
%word on that indexes
fprintf('No. of Non-zero elements in SPCs for 0.75 of cumalative variance is %f\n',k);
fprintf('List of words in 1st Sparse PC are \n')
words_list1=wordlist(ind_nonzero1)
```

Part A}

m =

  100

n =

   16242

Percentage Variance captured by first 3 PCs 0.190724
Variance by 1st PC 7345.246401

Part B}
No. of Non-zero elements in Sparse PCs for 0.75 of Variance in 1st PC is 18.000000
List of words in 1st Sparse PC are

words_list =

  1×18 cell array

  Columns 1 through 6

   {'case'}   {'computer'}   {'course'}   {'email'}   {'fact'}   {'god'}

  Columns 7 through 11

   {'government'}   {'help'}   {'number'}   {'power'}   {'problem'}

  Columns 12 through 16

   {'program'}   {'question'}   {'state'}   {'system'}   {'university'}

  Columns 17 through 18

   {'windows'}   {'world'}

PART C}

No. of Non-zero elements in SPCs for 0.75 of cumulative variance is 17.000000
List of words in Second sparse PC are

words_list1 =

  1×17 cell array

Columns 1 through 5

  {'bible'}   {'case'}   {'christian'}   {'course'}   {'email'}

Columns 6 through 10

  {'evidence'}   {'fact'}   {'god'}   {'government'}   {'help'}

Columns 11 through 15

  {'human'}   {'jesus'}   {'law'}   {'question'}   {'religion'}

Columns 16 through 17

  {'windows'}   {'world'}

So from the intuition by observing words we can say that-
1}The first Sparse PC holds mostly words related to topic "COMPUTERS".
 2}The 2nd Sparse PC holds mostly words related to topic "RELIGIONS AN POLITICS".

Question 2}}

```
load('autocomp.mat');
% Z=carbdata;
data=carbdata;
max_var=0;

%applying pca to data(witout mean shifting)
[U S V]=svd(data);
%total variance by pca
tot_var=sum(diag(S).^2)
[Fn,adj_v,cum_v]=sparsePCA(data,5,1,0,1);
```

```matlab
%percentage of variance
var_percent=cum_v(1)/tot_var;

fprintf('Without Mean Shifting\n')
fprintf('Cumulative Variance %f for %f non zero elements\n',cum_v,5);
fprintf('Percentage variance of the components are %f\n',var_percent*100);



%-----------------------------------------------------------------------
 data_mean=data-mean(data);
 [U S V]=svd(data_mean);
tot_var=sum(diag(S).^2)
[Fn,adj_v,cum_v]=sparsePCA(data_mean,5,1,0,1);
var_percent1=cum_v(1)/tot_var;
fprintf('With Mean Shifting\n')
fprintf('Cumulative Variance %f for %f non zero elements\n',cum_v,5);
fprintf('Percentage variance of the component are%f\n',var_percent1*100);
```

output

tot_var =

   3.5064e+05

Finished computing principal component number 1. #nonzeros=5
Without Mean Shifting
Cumulative Variance 255845.162313 for 5.000000 non zero elements
Percentage variance of the components are 72.965808

tot_var =

   0.5718

Finished computing principal component number 1. #nonzeros=5
With Mean Shifting
Cumulative Variance 0.249205 for 5.000000 non zero elements
Percentage variance of the component are43.580555