Joanna Masikowska

B9TB1710

CAPS10_B9TB1710(1).m CAPS10_B9TB1710(2).m

```
R1=dlmread("rating1.txt"," ",5,1);
R1=R1(1:5,:);

#decomposition into R1=P1*S1
[U1,W1,V1]=svd(R1,'econ');
P1=U1(:,1:3);
S1=W1(1:3,1:3)*V1'(1:3,:);

r6=[4,2,3]; #known ratings of 6th person
X1=[S1(:,1)';S1(:,3)';S1(:,7)']';
p6=r6*inv(X1);

#complementing data of 6th person
R1(6,1)=4;
R1(6,3)=2;
R1(6,7)=3;

for i=1:20
  if i!=[1,3,7]
    R1(6,i)=p6*S1(:,i);
  endif
endfor

Ra=[4,3,2,2,3,3,3,2,3,1,2,3,2,2,3,4,3,3,3];
result1=[R1(6,:);Ra;abs(R1(6,:)-Ra)]'

estim1=[result1(2,3)',result1(4:6,3)', result1(8:20,3)'];
mean1=mean(estim1);
med1=median(estim1);

printf("Median of differences between estimations \nand actual data is %f, and mean is %f.",med1,mean1);
```

I start from loading ratings data into matrix R1 (data of 5 persons) using function **dlmread** (I had problems with function **load**). Then I decompose R1 into U1, W1, V1 using singular value decomposition. I include function **'econ'** in order to eliminate rows of zeros in matrix W1. Then, I cut matrices U1,W1,V1 so that they are all of rank 3 but are still factors of R1. I create matrix P1 (size 5x3) from the shortened matrix U1, and matrix S1 (size 3x20) from the sum of shortened matrices W1 and $V1^T$.

Next, I want to predict unknown ratings of the new person. In order to do that, I want to find p_6 (size 1x3) in the following equation:

$$[R1(6,1) \quad R1(6,3) \quad R1(6,7)] = p_6 \times \begin{bmatrix} S1(1,1) & S1(2,1) & S1(3,1) \\ S1(1,3) & S1(2,3) & S1(3,3) \\ S1(1,7) & S1(2,7) & S1(3,7) \end{bmatrix}^T$$

I create matrix r_6 (size 1x3) which has only values of three known ratings, and matrix X1 which has columns 1,3,7 of matrix S1. Then, the above equation is equivalent to:

$$r_6 = p_6 \times X_1$$

By taking an inverse of X1 I get the value of p_6.

$$r_6 \times X_1^{-1} = p_6$$

Now, I can estimate ratings of any song by $R1(6,i) = p_6 \times S1(:,i)$.

I complete R1 with estimations. I repeat all the steps with data from rating2.txt. The only difference I need to adjust is the different size of matrix R2 (storing data from the file), which is 15x20.

CAPS10_B9TB1710(1).m CAPS10_B9TB1710(2).m

```
R2=dlmread("rating2.txt"," ",5,1);
R2=R2(1:15,:);

#decomposition into R2=P2*S2
[U2,W2,V2]=svd(R2, 'econ');
P2=U2(:,1:3);
S2=W2(1:3,1:3)*V2'(1:3,:);

r16=[4,2,3]; #known ratings of 16th person
X2=[S2(:,1)';S2(:,3)';S2(:,7)']';
p16=r16*inv(X2);

#complementing data of 16th person into R2
R2(16,1)=4;
R2(16,3)=2;
R2(16,7)=3;

for i=1:20
  if i!=[1,3,7]
    R2(16,i)=p16*S2(:,i);
  endif
endfor

Ra=[4,3,2,2,3,3,3,2,3,1,2,3,2,2,3,4,3,3,3,3];
result2=[R2(16,:);Ra;abs(R2(16,:)-Ra)]'

estim2=[result2(2,3)',result2(4:6,3)', result2(8:20,3)'];
mean2=mean(estim2);
med2=median(estim2);

printf("Median of differences between estimations \nand actual data is %f, and mean is %f.",med2,mean2);
```

Lastly, I compare my estimations with actual ratings of the new person in both cases. The results are as below.

コマンドウィンドウ

```
>> result1 =

   4.00000   4.00000   0.00000
   4.23347   3.00000   1.23347
   2.00000   2.00000   0.00000
   4.43954   2.00000   2.43954
   4.38326   3.00000   1.38326
   4.17720   3.00000   1.17720
   3.00000   3.00000   0.00000
   3.43954   2.00000   1.43954
   3.14979   3.00000   0.14979
   3.20607   1.00000   2.20607
   3.20607   2.00000   1.20607
   3.00000   3.00000   0.00000
   1.91632   2.00000   0.08368
   1.71025   2.00000   0.28975
   3.14979   3.00000   0.14979
   3.09352   4.00000   0.90648
   4.00000   3.00000   1.00000
   4.28975   3.00000   1.28975
   3.00000   3.00000   0.00000
   3.38326   3.00000   0.38326

Median of differences between estimations
and actual data is 1.000000, and mean is 0.902214.
```

コマンドウィンドウ

```
>> result2 =

   4.00000   4.00000   0.00000
   3.26521   3.00000   0.26521
   2.00000   2.00000   0.00000
   1.43091   2.00000   0.56909
   2.49526   3.00000   0.50474
   2.06190   3.00000   0.93810
   3.00000   3.00000   0.00000
   0.69249   2.00000   1.30751
   3.29798   3.00000   0.29798
   0.22492   1.00000   0.77508
   1.12459   2.00000   0.87541
   3.26909   3.00000   0.26909
   1.43751   2.00000   0.56249
   1.49979   2.00000   0.50021
   2.31549   3.00000   0.68451
   3.34886   4.00000   0.65114
   4.21450   3.00000   1.21450
   3.73124   3.00000   0.73124
   2.60018   3.00000   0.39982
   3.03495   3.00000   0.03495

Median of differences between estimations
and actual data is 0.569087, and mean is 0.622415.
```

1st column consists of estimated ratings of songs, where the number of a row stands for the number of a song. 2nd column is the actual rating of the corresponding song, and 3rd column has absolute values of differences between the actual and estimated rating. In rows where the 3rd column has zeros, the rating of a song was known prior to the estimation. Mean and median are calculated based only on estimated data.

I can see that when I had more data of people's ratings (result2), the mean and median of mistakes were smaller which means that my estimations where better in case 2.