Solutions, Homework 10

Assigned Thur/Fri, Mar 10/11 (Due Tues, March 29, 10PM), along with the iris data projection, previously assigned.

The Iris Data

This was from page 14 of the handout linked from Thursday, March 1st. To remind you what is was, here was the question:

> Revisiting the Iris Data: Go onto our class website and download the iris data again. Recall that the data is arranged as a 150×4 matrix, where there are 3 classes of iris- they are currently arranged in order.
>
> Here is the problem: Think of the data as 150 points in $\mathbb{R}^4$. Use the SVD of the mean subtracted data (as 150 vectors in $\mathbb{R}^4$) as basis vectors, and construct the two dimensional representation of the data (Recall our discussion of the low dimensional representation?).
>
> Before you set up the graph, you might recall that our classifier (from Homework 7) did a great job with Class 1, but got a little mixed up between classes 2 and 3. For your convenience, you might recall a confusion matrix (given).

SOLUTION: Here is the Matlab code to perform the projection to the low dimensional space.

```
% Plot the iris data in the ``best'' two dimensional space using the basis
% vectors from the SVD:

load IrisData
Xm=X-repmat(mean(X),150,1); % This mean-subtracts the data

%Do the reduced SVD on Xm to get basis vectors in R^4
%  Since Xm is 150 x 4, the basis vectors are the first cols of V

[U,S,V]=svd(Xm,'econ');  %Don't need the full SVD

Coords=Xm*V(:,1:2);  %Transposed because Xm is "tall", result is 150 x 2
Coords=Coords';      %Change it for the following plot command

%If the low dimensional data is in a 2x150 array called Coords, then:
plot(Coords(1,1:50),Coords(2,1:50),'r.');
hold on
plot(Coords(1,51:100),Coords(2,51:100),'b.');
plot(Coords(1,101:150),Coords(2,101:150),'m.');
hold off
```

Movie Data

The data represents a short movie of 109 frames, each where each frame is $120 \times 160 = 19200$ pixels. When you type `load author`, you should get an array $Y1$ that is 19200×109.

Goal: Compute the first two vectors for the column space of $Y1$, then project the data to it and plot the result. What is your estimate for the rank of $Y1$? Before you do the projection, we ought to mean subtract the data (given below).

See the original homework for sample code fragments- Below is the completed code that gives the solution

SOLUTION: Here is the code. For an estimate of the dimension, we typically use a number like 0.95 or 0.98. In this case, using 95% gives a dimension of 80. Using 98% gave a dimension of 90.

```
load author
```

```
Y1=double(Y1);
[mm,nn]=size(Y1);

% Find the mean of the movie frames, which is a point in R^{19200}, and
% mean subtract.  Show the mean in Figure 1.

   meanFrame=mean(Y1,2);
   Y1=Y1-repmat(meanFrame,1,nn);
   figure(1)
   imagesc(reshape(meanFrame,120,160));
   colormap(gray); axis square

% Find the Best Basis, and plot the first four basis vectors (later changed
% to 2 basis vectors).
[U,S,V]=svd(Y1,'econ');

% Optional plots:
figure(2)
for j=1:4
    subplot(2,2,j)
    imagesc(reshape(U(:,j),120,160));
    colormap(gray); axis square
end

% The low dimensional representation:
Coords=U(:,1:2)'*Y1;

figure(3)
plot(Coords(1,:),Coords(2,:));

% Estimate the rank:  We will normalize the singular values,
%  then see what rank we need to keep about 95 percent of the "energy" of
%  the system.

S1=diag(S)./sum(diag(S));
S2=cumsum(S1);
for j=1:length(S2)
    if S2(j)>=0.95
        fprintf('Dimension is %d\n',j);
        break
    end
end
```