

CS340 Winter 2010: HW2
Out Mon. 24th January, due Wed. 2nd
February

1 PCA Image Visualization

In this homework, you need to implement the image visualization PCA tool discussed in class. Your task is to modify the following piece of code. Whenever you encounter a ??? sign, replace it with the appropriate code.

```
load twosfours;
X = data';
% X has 30 columns corresponding to 30 16x16 images.
% The first 20 columns correspond to images of 2s and the last
% 10 columns to images of 4's

[d N] = size(???);

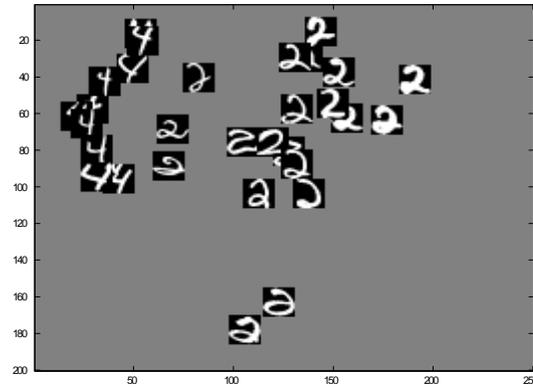
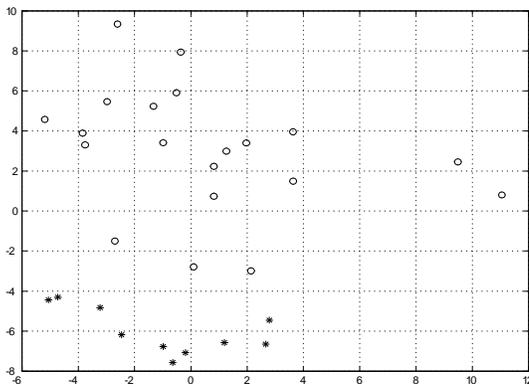
% Compute mean of each row of X and center data by subtracting
each row mean
XMean = mean(???)'; % vector of size d*1, compute mean for each
row
Xc=X-repmat(???,1,N);
% Find eigendecomposition
[W D]=eig(???);
% compute principal components associated to the two largest eigenvectors
WReduced=W(:,???);
pc=WReduced'*Xc;

figure(1)
% Plot principal components.
```

```

clf;
plot(pc(1,1:20),pc(2,1:20),'or',???,???,'*b');
grid;
print -deps pcavisual1
% Plot images on a background of size 120 by 250.
I= zeros(200,250);
% I is the background image on which small images are placed.
% now we scale the components to be integers for indexing:
pc(1,:)=pc(1,:)+6*ones(1,N);
pc(2,:)=pc(2,:)+9*ones(1,N);
pc = round(10*pc);
% now we add the individual images to the background image:
for i=1:N
    I(pc(1,i):pc(1,i)+15,pc(2,i):pc(2,i)+15) = reshape(???,???,16)';
end
figure(2)
% Visualization:
clf;
colormap('gray')
imagesc(I)
print -deps pcavisual2
Plot your answers. They should look like this.

```



Hand in the completed code.

2 Image Compression

Repeat the image compression example presented in class. However, this time you should use your own photo. Initially you should use any image package to generate a jpg colour image of yourself that is 512 by 512 pixels. Then load this image to matlab using the command `imread`. Convert the image from rgb colour space to gray-scale using `rgb2gray`. Convert the gray-scale image to double precision using the command `double`. To compute the PCA, I want you to use the matlab function `eig`. Important: always center the data before applying PCA (as in Question 1) but do not rescale them. Hand in all your code.

1. In the first application of the PCA, use the columns of the image as data points. Generate a figure similar to figure 1 (except it should be your own picture!) which displays \mathbf{X} in the upper left corner and then $\widehat{\mathbf{X}}$ for $k = 1, 5, 10, 20, 25, 30, 35, 40$ (Note: the matlab command `subplot` will be useful). What is the compression gain $\frac{dN}{dk+kN}$ achieved as a function of k ? What is the average squared reconstruction error $\frac{1}{N} \left\| \mathbf{X} - \widehat{\mathbf{X}} \right\|^2$ as a function of k ?
2. In the second application of the PCA, partition the image in non-overlapping smaller images of 8×8 pixels and use these small images as data points. (Note: in matlab a matrix \mathbf{A} is converted into a column vector \mathbf{a} by means of the command `a = A(:)`. You can reconstruct \mathbf{A} from \mathbf{a} by means of the command `reshape`). Generate a figure similar to figure 2 (except it should be your own picture!) which displays the original image in the upper left corner and then its approximation obtained for $k = 1, 2, 3, 4, 5, 6, 7, 8$. What is the compression gain achieved as a function of k ? What is the average squared reconstruction error $\frac{1}{N} \left\| \mathbf{X} - \widehat{\mathbf{X}} \right\|^2$ as a function of k ?
3. We have seen in class that when $d \gg N$ an alternative approach to compute the eigenvectors \mathbf{u}_i of $\widehat{\Sigma} = \frac{1}{N} \mathbf{X} \mathbf{X}^T$ consists of computing the eigenvectors \mathbf{v}_i of the Gram matrix $\mathbf{X}^T \mathbf{X}$. Prove that if \mathbf{v}_i is an eigenvector of $\mathbf{X}^T \mathbf{X}$ associated to the eigenvalue λ_i such that $\|\mathbf{v}_i\| = 1$ then $\mathbf{u}_i = \lambda_i^{-1/2} (\mathbf{X} \mathbf{v}_i)$ is an eigenvector of $\widehat{\Sigma}$ associated to the eigenvalue λ_i/N and that it satisfies $\|\mathbf{u}_i\| = 1$. Use this result to



Figure 1: Original picture and its PCA approximations when data correspond to columns.

repeat the previous question when the image is now partitioned in non-overlapping smaller images of 64×64 pixels and use these small images as data points. Compare visually the reconstructed image for $k = 1$ obtained in this question (64×64 pixels) and in the previous question (8×8 pixels). Could you provide an intuitive explanation for these results?

3 Text Retrieval

This exercise teaches you how to build a simple text search engine and text compression scheme. The file `matrix.dat` contains a matrix of documents by words for the $N = 9$ documents that appear in the file `documents.pdf`. The list of the $d = 460$ words (in order) appears in the file `literals.dat`; that is we have $\mathbf{X} = \text{matrix}^T$ using the notation used in the lectures. Center the data before applying the PCA (as in Question 1) but do not rescale them.

1. Compute and plot the eigenvalues associated to $\hat{\Sigma} = \frac{1}{N} \mathbf{X}_c \mathbf{X}_c^T$ where \mathbf{X}_c is obtained by centering the data.



Figure 2: Original picture and its PCA approximations when data correspond to non-overlapping blocks.

2. Compute the PCA approximation $\hat{\mathbf{X}}$ of \mathbf{X} using $k = 3$. Create a plot with 2 subplots using the matlab command `subplot`. In the first subplot do `image(X)`. In the second subplot, plot the image $\hat{\mathbf{X}}$. What do you conclude from comparing the two plots? What is the compression gain? what is the average squared reconstruction error $\frac{1}{N} \left\| \mathbf{X} - \hat{\mathbf{X}} \right\|^2$?
3. What are the angles, in degrees, between the first document projected in 3D (that is the approximation $\hat{\mathbf{x}}_1$ of \mathbf{x}_1 for $k = 3$) and the 8 remaining projected documents? By looking at the documents, can you explain these results?

4 Probability: Conditioning and marginalisation:

A band called Radiohead is inspired by an old band called The Beatles. 50% of music critics think the beatles was a great (G) band, 40% that it

was moderate (M) and 10% that it was awful (A). These critics have also compiled the following table:

		B_2		
		G	M	A
B_1	G	0.8	0.1	0.1
	M	0.1	0.9	0
	A	0.2	0.3	0.5

The table says that the probability of a new band (B_2) being great given that the inspiring band (B_1) was great is $P(B_2 = G | B_1 = G) = 0.8$. Similarly, $P(B_2 = G | B_1 = M) = 0.1$, $P(B_2 = M | B_1 = A) = 0.3$, and so on. Note that the numbers in the rows add up to 1, so the table is a probability transition matrix.

1. Given what the critics think of the Beatles and the fact that the Beatles inspired Radiohead, what is the probability that Radiohead is a great band?
2. What is $P(B_1 = G | B_2 = G)$?