HW 7

Greg Dehmlow, Colin Murphy, Nate Holland

May 5, 2014

Instructions: unzip the directory and go into it. Run the hw7part1 and hw7part2 scripts to generate images and whatnot.

1 Warmup

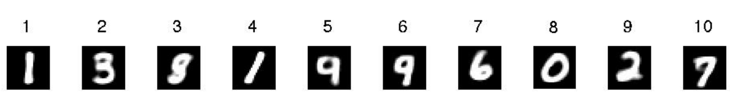
Here's what the centroids initially looked like for k = 5:

Yikes. The above implementation used super naïve initialization - we found the minimum and maximum values across all dimensions and drew k random samples uniformly from that hypercube. We also weren't normalizing the inputs in any way.

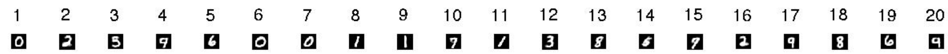
This obviously was not a very good approach. So we switched to using a slightly less naïve initialization - randomly sampling k points from the data set as our initial centroids. This went much much faster. We also decided to properly normalize the mean and standard deviation of the data. With the extra time saved, we decided to iterate, say, 10 times, keeping track of the best performing clustering. What is best performance in this case? There are many ways that this could be approached; after doing some research on the internet, one interesting approach was to, for each cluster, calculate the mean distance from the centroid to each point belonging in the cluster. So during k-means, we would get k mean distances. We would optimize by minimizing this variance because we want each cluster to be roughly the same size.

Using the new approach, on the MNIST 10,000 digit set, here's what centroids for k=5 look like:

k=10:

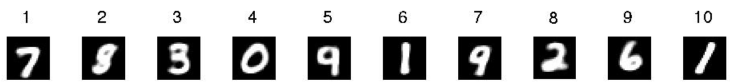


k=20:



Overall, it looks pretty good. The differences in appearances between the different number of k's makes sense: k=5 has a lot more blurring since different digits have to fit in the same centroid. For k=20, some boxes will overfit and as a result multiple centroids will belong to the same digit. For k=10, there's a little bit of duplication, but most digits are represented clearly.

Since we have the labels, we can (sort of) verify how good the clusterings are, although this kind of entirely defeats the point. But it's a fun exercise none-the-less. For each cluster, we see what labels the points that belong to it have. We can then assign the probability that a point assigned to a cluster is the same as the rest in the cluster. The larger the max probability for each cluster, the lower the error. In this method, I iterate over 10-means say, 10 times, the same way as above, except I quantify this new error each time and keep track of the centroids which minimize this error. The result is seen below, and the 'error' was 33.6%:



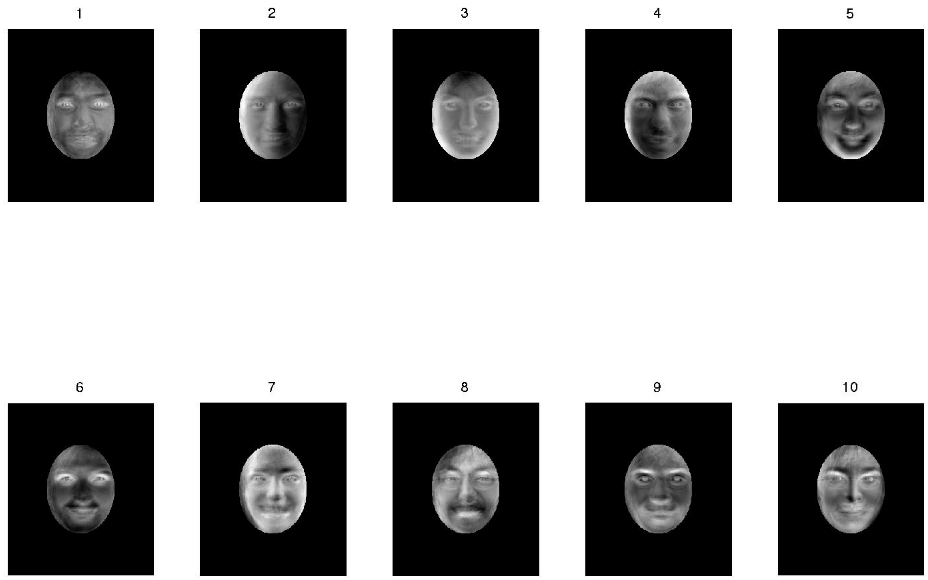
Perhaps not surprisingly, it isn't that much better. Perhaps we should consider a different performance criteria during our iterations. 4 is still conspicuously absent (probably merged with the 9s) and 5 seems to be getting along quite nicely with 3 and 8. What is also interesting about having the labels is that we can use this method see which digits tend to get mixed up the most. If a cluster is less than 70% 'sure' about its label, it can print other labels that have at least 10% of the data points with that index. So for this set of centroids:

Our output was:
4 [1], (0.37) looks like: 7 (0.21), 9 (0.33),
7 [6], (0.38) looks like: 4 (0.18), 9 (0.24),
0 [7], (0.62) looks like: 5 (0.19),
3 [9], (0.44) looks like: 5 (0.29), 8 (0.18),
3 [10], (0.38) looks like: 5 (0.15), 8 (0.38)

As was suspected, 4, 9, and 7 (centroids 1 and 6) get confused, as do 3, 5, and 8 (centroids 9 and 10).

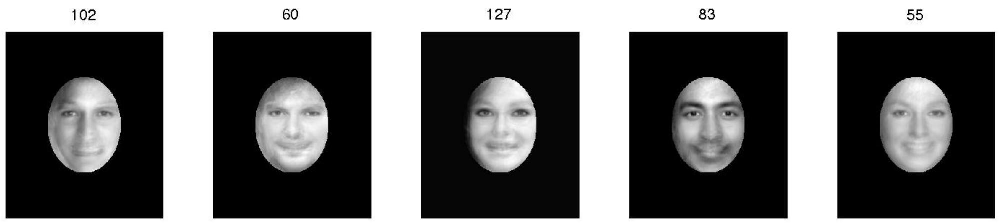
2 Eigenfaces

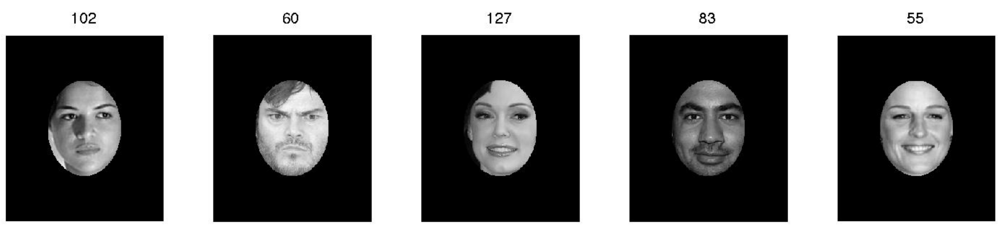
2.1 What kind of variations do the top eigenfaces seem to correspond to?

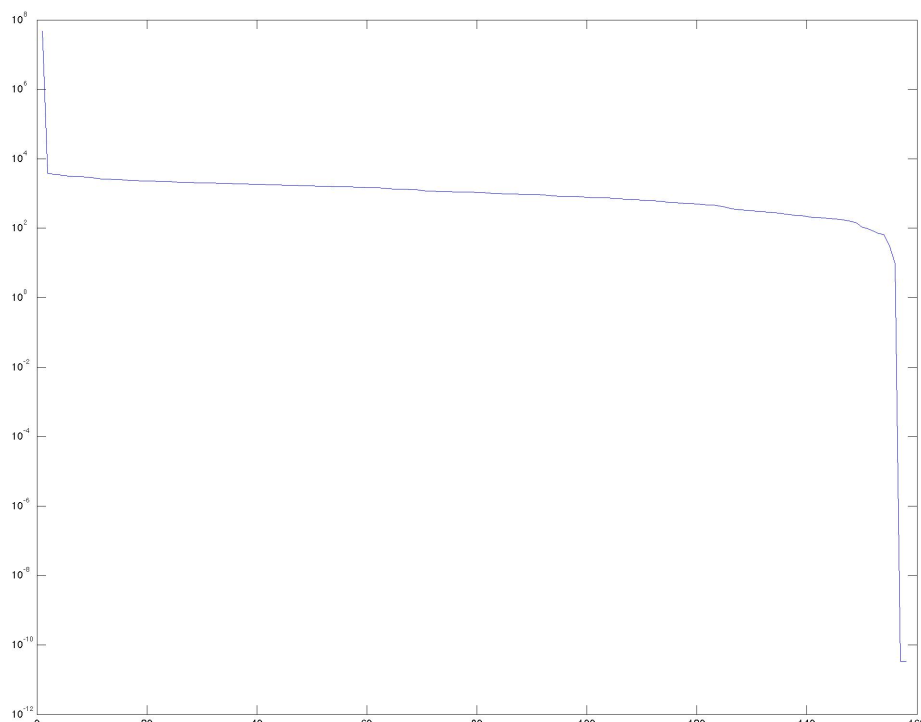


These eigenfaces seem to roughly correspond to different lighting conditions, most obviously diferent light source directions, as well as emphasizing different aspects of the face: lighter eyes, eyeshadow, lips, and nose. Some eigenfaces seem more smiley, like 7, whereas 10 seems feminine.

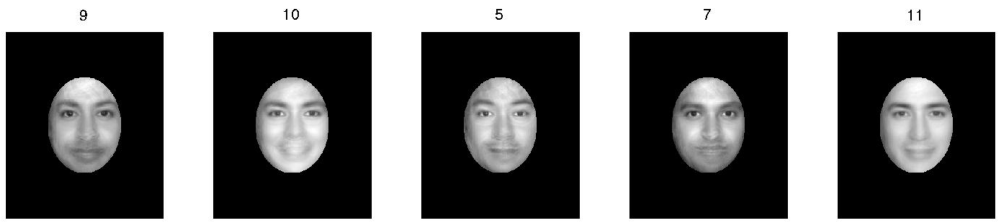
2.2 L2 error of 5 celebrity faces reconstructed using top 10 eigenfaces

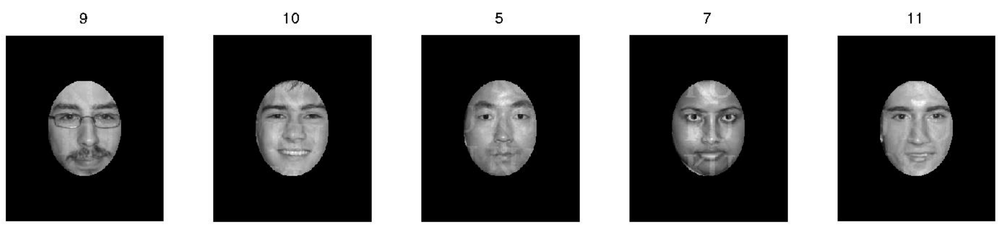


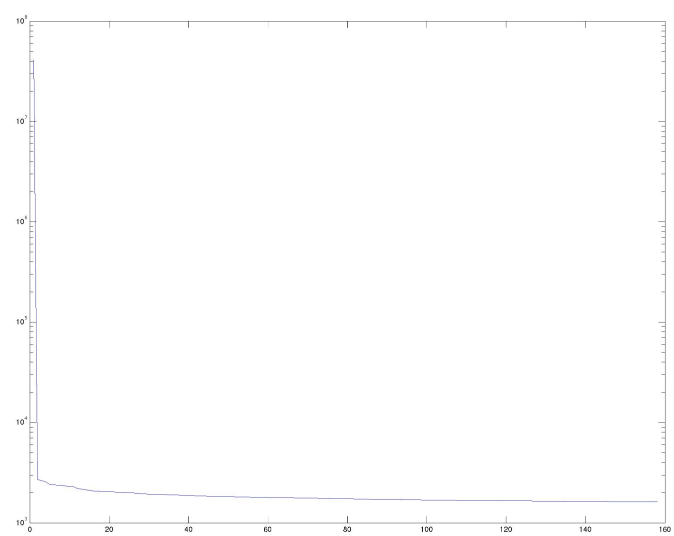




2.3 L2 error of 5 student faces reconstructed using top 20 eigenfaces







The most interesting thing to notice here is that the celebrity eigenfaces quite good at reconstructing student faces. The difference in the errors is that each celebrity face can be completely reconstructed by using all eigenfaces, which is why there's the sharp drop in error for celebrity eigenfaces close to the number of faces in the database. The student L2 error, on the other hand, levels off pretty quickly and stays there.

3 SVD

Part one

$A = \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix}$

$A^T A = \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$

$\lambda_1 = 8, \lambda_2 = 2 \Rightarrow \sigma_1 = 2\sqrt{2}, \sigma_2 = \sqrt{2}$

$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$

$\lambda = 8$

$A^T A - \lambda I = \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \Rightarrow x_1 - x_2 = 0 \Rightarrow x_1 = x_2 \Rightarrow v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

$\lambda = 2$

$A^T A - \lambda I = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow x_1 + x_2 = 0 \Rightarrow x_1 = -x_2 \Rightarrow v_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

$V = \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

$Av_1 = \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix} = \sigma_1 u_1 \Rightarrow u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$Av_2 = \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ -\sqrt{2} \end{bmatrix} = \sigma_1 u_1 \Rightarrow u_1 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$U = \begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

$U\Sigma V^T = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix} = A$

Part 2

$A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$

$A^T A = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix}$

$\lambda_1 = 10, \lambda_2 = 0 \Rightarrow \sigma_1 = \sqrt{10}, \sigma_2 = 0$

$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{bmatrix}$

$\lambda = 10$

$A^T A - \lambda I = \begin{bmatrix} -5 & 5 \\ 5 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \Rightarrow x_1 - x_2 = 0 \Rightarrow x_1 = x_2 \Rightarrow v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

$\lambda = 0$

$A^T A - \lambda I = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow x_1 + x_2 = 0 \Rightarrow x_1 = -x_2 \Rightarrow v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$

$V = \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

$Av_1 = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} \\ \sqrt{2} \end{bmatrix} = \sigma_1 u_1 \Rightarrow u_1 = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$

$Av_2 = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \sigma_1 u_1 \Rightarrow u_1 = \begin{bmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$

$U = \begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}$

$$U\Sigma V^T = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \sqrt{5} & \sqrt{5} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} = A$$