Data Experiments with the Best Basis
Modeling II

1. What is the effect of mean subtraction? Construct a one dimensional set of data of 100 points in $\mathbb{R}^2$: $2 \le x \le 4$, $y = 3x + 2$.

   (a) Find the best basis before doing any mean subtraction. In particular, note the values in U, Σ, V.

   (b) Mean subtract (mean from $\mathbb{R}^2$) and find the best basis.

   (c) Mean subtract (mean from $\mathbb{R}^{100}$) and note the singular values.

   (d) Double mean subtract, and again find the best basis.

   Plot the four data sets (on the same axis) and answer the question: What is the effect on the basis of mean subtraction? Should we mean subtract?

2. What is the meaning of the singular values (or eigenvalues of the covariance)? Let us construct a matrix A with some desired properties using the SVD. Let A be a 2×2 matrix so that the basis vectors for the domain are the standard vectors. Let the basis vectors for the range be the standard basis vectors rotated by 30 degrees. To get these vectors,

   ```
   V=eye(2);  %Domain Vectors
   U=[cos(pi/6) -sin(pi/6); sin(pi/6) cos(pi/6)]; %Range vectors
   ```

   To see how A maps the domain to the range, we will determine what happens to 100 points on the unit circle. To get these points,

   ```
   t=linspace(0, 2*pi);
   X=[cos(t); sin(t)];  %X is 2 x 100
   ```

   For each set of singular values below, compute A, AX, and plot the resulting points:

   (a) Singular values: $1, 1$

   (b) Singular values: $3, 1/2$

   (c) Singular values: $1, 0$.

   (d) Singular values:: $2, 3$

   How does changing the singular values change the range of the unit circle? Use this information to determine the value of the following quantity for A found in (a)-(d):

   $$\max_{x \neq 0} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

For the eigenfaces, we'll want to use Matlab to do some things to pictures. In the following, we assume a photo is 262×294. Note that $262 \cdot 294 = 77028$. The variable `Face` will be a data array, and `x` will be a vector.

- To change `Face` to a column vector $\mathbf{x}$: `x=Face(:)`
- To change $\mathbf{x}$ back into an image: `Face=reshape(x,262, 294)`
- To visualize `Face`: `imagesc(Face); colormap(gray)`
- Given an $n \times k$ matrix U (k columns in $\mathbb{R}^n$), and a data point $\mathbf{x} \in \mathbb{R}^n$, $[\mathbf{x}]_u = U^T\mathbf{x}$. This is also the "low dimensional representation" of $\mathbf{x}$, which has k values. Note that if the matrix X is $n \times p$ (representing p data points in $\mathbb{R}^n$) we can compute all low dimensional representations at once: $U^T X$ (which is $k \times p$).
- Given that $\mathbf{y} \in \mathbb{R}^k$ is the low dimensional representation of $\mathbf{x}$, the representation back in $\mathbb{R}^n$ is: $U\mathbf{y}$, or in terms of $\mathbf{x}$, $UU^T\mathbf{x}$, which is also the projection of $\mathbf{x}$ into the columnspace of U.
- Given $X = U\Sigma V^T$, and $\hat{U}, \hat{\Sigma}, \hat{V}$ are $n \times k$, $k \times k$, $p \times k$ respectively for any k, then
$$\hat{U}\hat{U}^T X = \hat{U}\hat{\Sigma}\hat{V}^T$$
is the $k-$dimensional SVD reconstruction of X (like we did with the clown image, where we had `U(:,1:k)*S(1:k,1:k)*V(:,1:k)'`).

What we will do with the faces:

1. Load the faces data and create a matrix 77028×30.
2. Find the mean face and plot it.
3. Mean subtract the data matrix.
4. Compute the best basis for the space of these faces. Plot the first six basis vectors as faces.

   BEFORE YOU DO THIS: What if the computer can't factor this big of a matrix? Remember that the vectors in $\mathbb{R}^{30}$ and in $\mathbb{R}^{77028}$ are related!
5. Compute the two dimensional representation of the faces (the data will be 2×30 or 30×2 depending on how you do it). Plot these in $\mathbb{R}^2$.
6. Pick a face to see the partial reconstructions as we add more and more basis vectors. Plot these as faces (remember to add the mean back in!) using the `subplot(2,3,k)` command.