STA 5107/4013: Home Assignment # 1

Spring 2021/Due Date: January 21

1. **Principal Component Analysis**: For each of the datasets provided to you, perform the following operations:

   - If $n \leq 3$, display the data as a scatter plot.
   - Perform PCA of the given data, use $d = 2$ ($k = 100$ in all the examples).
   - If $n \leq 3$, display the top two principal directions of variability overlaid on the scatter plot of the data.
   - Project the original data to the first two principal dimensions, i.e. $d = 2$, using $Z = U^T X$, where $U = V(:, 1:2)$, and display Z as a 2D scatter plot. (Note, if $n = d$ then Z is simply a rotation of X).
   - Plot the singular values of the covariance of X. How many singular values are sufficient to obtain 95% of the total variance.

   The data sets are provided as matrices $X \in \mathbb{R}^{n \times k}$, k points in $\mathbb{R}^n$.

2. **PCA of Face Dataset**: For the image dataset provided to you, perform its PCA and display the following results:

   - Image of the sample mean face
   - Plot of singular values of the covariance matrix
   - Images of the first three principal eigenvectors of the covariance matrix.
   - Images of any three (arbitrary) faces, their reconstructions using $d = 20$ components, and the absolute difference of the error.

   The dataset is provided to you as a mat file with variable $X \in \mathbb{R}^{3584 \times 225}$. Each column denotes a face image of size 64×56. You use the `reshape` command in matlab to convert a large vector into an image — $I = reshape(X(:, 1), 64, 56);$.