Statistics 321, Statistical Computing, Home Work 9

Handed Out: Oct. 31, 1996 Due: Nov. 7, 1996

The goal of this is to investigate the "nonrobustness" of principle components. The main idea is that the direction of the eigenvectors can be driven by a very few "outliers".

1. Write a Matlab script which generates 100 two dimensional pseudo random vectors from the $N(\mu, \Sigma)$ distribution with
$$\mu = \begin{pmatrix} -6 \\ -4 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 2 & -1.4 \\ -1.4 & 1 \end{pmatrix}.$$
Turn in a scatterplot of the data, with lines overlaid, which show the directions of the principle component eigenvectors.

2. Start with the same pseudo data as in part 1 above, and add the single new data vector $\begin{pmatrix} 14 \\ 19 \end{pmatrix}$. Again make a scatterplot, with the new principle component axes overlaid. Comment on the difference between the principle component directions here and in part 1 above.

3. Construct your own example, this time in 3-d, which is a mixture of three normal distributions, which shows that when the data are multimodal, the principle component directions can be quite different from those of the Normal subpopulations. Visualize the result with a movie. "Turn in" only a matlab script which runs the movie (by email, using the PINE "attach" feature please), which I will then run. Bonus points will be awarded for especially creative movies, and/or figuring out how to get principle component directions to overlay rotating point clouds)

Notes:

1. To save coding time, borrow code from the example script: `s322eg11.m`
2. To "use the full page while printing", check out the command `orient`.
3. For parts 1 and 2, to make the eigenvector directions really look perpendicular, check out the command `axis`.