HW3 solutions to questions 7,8,9)

7 The first subplot shows the results of k-means clustering. It does as expected and divides the data by a line (not what we would like). The rest of the subplots show the stages of the spectral clustering algorithm. The second subplot gives the affinity matrix with the diagonal zeroed. The third subplot gives the normalized Laplacian matrix (which is just the affinity matrix (with zero'd diag) normalized (each entry a_{ij} is divided by the product $\sqrt{d_i}\sqrt{d_j}$ where $d_i = \sum_j a_{ij}$. The third subplot shows the top two sorted eigenvectors (top eigenvector on the left). The fourth subplot shows the same thing with the rows normalized (so that they lie on a unit sphere). The fifth plot shows the final result of the spectral clustering algorithm after the final kmeans step. Spectral clustering gives the desired result.

8) Sigsq sets the scale for similarity. If sigsq is too small then most affinity values are very small. If sigsq is too large then many affinity values are close to 1. There is a small range of sigsq that "works" for the given dataset.

9) There is not a sigsq that works for all the random datasets generated with the same code (from the same pdf). If there are large gaps in the rings then spectral clustering will tend to split the ring at the large gap and cluster part of the big ring with the small ring. This question was designed to show you how it is not easy to find an optimal sigsq and to show you how sensitive the algorithm is to the exact data sample and sigsq.