

ASTR 600, Homework #6
Chapter 9-- Clustering

(I will give you the data set for this homework problem)

1. First, make a scatterplot of the data. Describe what you see. Just by looking at the scatterplot, how many clusters do you predict? How do you know what is noise and what is data?
2. Make a Cluster Dendrogram out of the data. Describe what it tells you.
3. Now, divide your tree into 5 groups. Plot X vs Y, using a different shape or color for each group. What can you conclude? Do you agree with this clustering pattern? Why or why not?
4. I had you break the data into 5 clusters. But YOU might not see 5 clusters. Choose a different number of clusters that you think is appropriate. Using k-means cluster analysis, re-group your data. Plot it the same way that you did for #3. Do you think it is a better or worse clustering, and why?
5. Use a density-based cluster analysis to cluster the data. (Hint: load the package 'fpc'.) Use it to find the clusters and the noise. Start with a "reach" of 1, and a "minimum number of points" of 10. Then, vary the reach and points until you believe it most accurately clusters the data. Then, plot your final clustering solution with the clusters and the noise clearly identified.
6. Use the friends-of-friends algorithm to cluster the data. (You might want to refer to <http://homes.cs.washington.edu/~magda/papers/kwon-ssdbm10.pdf>) Then, plot your final clustering solution with the clusters and the noise clearly identified. Also, please explain your code as you go.