Data Mining II
Homework 2

**ISLR – Introduction to Statistical Learning

1) (10 points) ISLR text: Chapter 10 Question 9

2) (10 points) ISLR text: Chapter 10 Question 11

3) (10 points) Access the data "seeds data" (on UB learns).
This data contains the geometrical properties of kernels belonging to three different varieties of wheat (seed group). The original data can be found: https://archive.ics.uci.edu/ml/datasets/seeds, although I have modified the data slightly.
   a) Cluster the data based on single-linkage, average linkage, and complete-linkage agglomerative hierarchical clustering. Do not use the "seed group" column to perform the clustering, but use it to help evaluate your results.

      Decide on the groupings, and justify it, for all three methods. The justification should be based on a measure (you select which) that we learned in class.

      Which method "performed" the best and which method performed the worst? Was the result in line with your expectations?

   b) Cluster the data based on K-means or K-medoids. Use an analytical technique to justify your choice in "k". How did the performance compare to the hierarchical clustering of part a? Which did you feel was a better method for this data?