1. Implement K-means using the greedy algorithm described in the handout, using random assignment to determine the initial mapping $\phi$. Also implement a global solution to the K-means problem by enumerating all partitions. Choose a sample size $n$ and partition number $k$ that are tractable for the global search, and carry out a simulation study to determine how similar the greedy solution is to the global solution. You will need to make reasonable decisions as to how the data are simulated, how you assess the discrepancy between two solutions, and how the simulation results are summarized.

2. Suppose you are given the solution to a hierarchical clustering problem in the form of a matrix $Q$ where $Q_{ij}$ is an integer label for the cluster to which object $i$ belongs in partition $j$ (i.e. $X_i$ and $X_k$ are in the same component of partition $j$ if and only if $Q_{ij} = Q_{kj}$). Determine a reordering of the objects (i.e. a permutation) so that the permuted objects are consistent with $Q$ in the following sense: for each partition, the objects in a given component of the partition are contiguous. This is the first step toward writing a program to draw the dendrogram (but you only need to find the permutation).

3. Suppose we have a $2 \times 2$ table with independent rows and columns, where the row probabilities are $p$ and $1 - p$, and the column probabilities are $(1 - p)^2$ and $1 - (1 - p)^2$. Use a numerical method to find the MLE for $p$ when the data are

\[
\begin{array}{cc}
132 & 90 \\
430 & 348 \\
\end{array}
\]

and when the data are

\[
\begin{array}{cc}
13 & 12 \\
44 & 31 \\
\end{array}
\]

Note that you do not have to guarantee that you have found the global MLE, just a local solution will suffice.