HW #5 (week 6)

- Due date: Fri, July 24 by 11:59pm EST. If you need extra time, just let me know.
- Download the data files to your local directory
- Send me both your script for analyzing the data (lastname_firstname_hw5.m) and the figures you make as .png files.
- Turn in the assignment by email at mlnguyen@princeton.edu

In this HW assignment, you will apply both k-means and hierarchical clustering to fMRI functional connectivity data. So far, we've looked at the neural response over time (a "timeseries" or "timecourse") in individual areas of the brain. For example, in any earlier assignment, you plotted the timeseries of three brain regions during a theory of mind task, which showed that these regions have greater activity during tasks that require mentalizing than tasks that do not.

fMRI is increasingly taking a network approach to the brain. Rather than looking at the response in individual brain regions, network-based approaches instead examine how activity in groups of regions covary (or don't) during tasks (or even at rest). Covariance among regions is often measured with functional connectivity, which is just a fancy word for saying correlations among regions. For example, activity in different regions of visual cortex are very strongly correlated, suggesting these regions are all involved in similar cognitive tasks (e.g. perception).

By computing functional connectivity among different brain regions covering the entire brain and then applying clustering methods, researchers have defined a small number of functional networks in the brain. A "network" in this sense is just a set of regions that have similar patterns of activity. Here, you will calculate functional connectivity among a set of 61 ROIs (regions of interest, some bit of the brain) defined from my recent teaching paper (you're all familiar with this by now) and then apply clustering methods to derive networks. We'll then look at whether the networks derived from this dataset are similar to the networks defined on a resting state data set.

Problem 1: functional connectivity

a. Load the `mean_roi_data.mat` file into your workspace. This file contains three variables:
    i. `mean_roi_tc` contains the mean timecourse, averaged across 20 subjects, for each of 61 regions-of-interest (61) which cover regions of the brain that respond reliably to temporally-extended stimuli. The data matrix is nROIs x time, so row m is the average timecourse for ROI m.
    ii. `roi_names` contain the names of the 61 ROIs are in the same order as mean_roi_tc
    iii. `roi_networks` contains the network assignment of each of the 61 ROIs based on previous clustering on resting state data. There are 6 networks: Visual, Auditory-

Language, Attention, Executive, Default Mode Network 1, and Default Mode Network 2.

b. Do the functional connectivity analysis: correlate the timeseries of each ROI with every other ROI. In the end, you should have a 61 x 61 matrix with the r-values for each of these correlations such that the r-value at row = 1 and col = 2 is the correlation of ROI 1 with ROI 2, the r-value at row = 32 and col = 15 is the correlation of ROI 32 with ROI 15, and so on. The matrix should be symmetric and the diagonals should be 1s (the correlation of each ROI with itself is 1).
c. Using `imagesc(mat)` to visualize the results of your functional connectivity analysis. Add the `roi_network` labels as x-tick or y-tick labels to get a sense of the similarity of responses within and across networks. Use `xtickangle(angle)` to rotate the labels for readability.

Problem 2: k-means clustering

In this problem, you'll run k-means clustering on the functional connectivity matrix you made in Problem 1.

a. Run k-means clustering with `k = 1:10` clusters with the `kmeans` function. Be sure to store the cluster indices, the coordinates of each cluster's centroid, and the sum of squared differences for each k in a variable.
b. Make an elbow plot using the average sum of squared differences for each k from 2a. Use this plot to select the number of clusters.
c. Using the number of clusters selected in 2b, re-arrange the functional connectivity matrix so that ROIs from the same cluster are next to each other. Make a new visualization with `imagesc` with the reordered functional connectivity matrix. You should see "blocks" of similar colors, indicating your clusters. Add a `colorbar` and `xtick` labels with the ROI network names. Save this figure. What do you observe?

Problem 3: agglomerative hierarchical clustering

In this problem, you'll run hierarchical clustering on the functional connectivity matrix you made in Problem 1.

a. Calculate the Euclidean between regions using the `pdist` function. Pass the entire connectivity matrix to this function.
b. Run agglomerative hierarchal clustering on the distances using the `linkage` function. Set the measure of distance to average.
c. Visualize the results using the dendrogram function. Set the second argument to 0 to see every ROI on the dendrogram. Add the `roi_network` labels as `xtick` labels. Save this figure How do the clusters differ from k-means clustering?
d. For your own edification, repeat 3b and 3c using single, complete, and centroid measures of distance. Notice that the selection of distance measure can have significant changes on the dendrogram.