BME 295: Analysis of Neural Time Series Data (Spring 2018)
Homework 1: Plotting & Preprocessing

For the first several homework assignments, we will analyze EEG data from a P300 experiment performed by Ulrich Hoffman, Jean-Marc Vesin, Touradj Ebrahimi, and Karen Diserens in Lausanne, Switzerland. The complete dataset is freely available online at http://mmspg.epfl.ch/BCI_datasets; however, all of the data and information you will need for the assignment will be available on the course website.

To turn in this assignment, please upload a single document (.doc, .pdf) to the EEE website. This should include answers to all of the questions, as well as any figures or code you generate (I have highlighted these in gray). For this assignment, you will be writing two functions, and I would recommend putting the rest of the code (to call the functions and generate the figures) in a single script called "Homework1.m" or something of that nature. Please copy/paste the two functions and the script at the end of your submission.

NOTE: To add MATLAB figures to a Word document, choose "Copy Figure" in the "Edit" menu of the figure, and then paste it into Word.

Download and unzip the data from the course website. The folder contains the following:

- Two pdf files, "Hoffman07_P300" and "P300_DescriptionOfData" that describe the data.
- Six Matlab .mat data files, which contain EEG data from the P300 experiments. These have names like "sub8_sess4_1."
- A data file called "channel_locations" which describes the locations of EEG electrodes for a 32-channel 10-20 montage in 3-D space.
- A function from the EEGLab toolbox called "topoplot". You can use topoplot to create a topographical plot of EEG data. Note that this function will call finputcheck.m, icadefs.m, and readlocs.m, so you'll likely need all of them.

With those details out of the way, we're ready to get started! First, let's get to know the data:

1. Open the manuscript that was written using the P300 dataset (Hoffman07_P300.pdf).
    a. What is a "P300"?
    b. How is the P300 utilized in Brain-Computer Interface devices?
    c. How many subjects were included in the study, and what two groups were being tested?
    d. How many channels of EEG were recorded?
    e. In a few sentences, describe the task performed by the subject in the first session of the experiment.
    f. How long did each stimulus (image) remain visible, and how much time elapsed between successive images?
    g. How did sessions 2, 3, and 4, differ from session 1?

2. Open the PDF that contains a description of the data (P300_DescriptionOfData.pdf)
    a. How many rows (channels) does the `data` matrix have? Note that this is different than your answer for 1.d. The last two electrodes are placed over the mastoid processes (behind the ears) and will not contain neural data.
    ** A picture of the electrode locations is shown in Figure 2 of the manuscript. **
    b. What is the sampling frequency of the data?
    c. At what time index does the first image appear? (In other words, to access the EEG data at the appearance of the first image, we can use the following MATLAB command: `data(:,x)`. What is `x`?)

Now let's use MATLAB to take a look at the data.

3. Load the dataset sub8_sess4_6.
    a. Plot each channel in the `data` matrix on the same figure. The x-axis should be time (in seconds). Recall that you can do this with a single command; if you use a matrix as an input to the plot function, each column will be plotted in a different color.

    You should notice two things: (1) each channel has a different DC offset and (2) all of the signals have extra zeros at the end of the session.

    b. Write a small piece of code to remove the extra zeros and then subtract the mean of each channel. Note that the zeros begin at a different time in each dataset, so you'll need to devise some logic to identify this time index. Plot the data again to verify that you have done this correctly.

4. Write a function called "`extractAllTrials`" that will extract all trials for a single channel:

```
[trials, time] = extractAllTrials(data, events, elec, t1, t2);
```

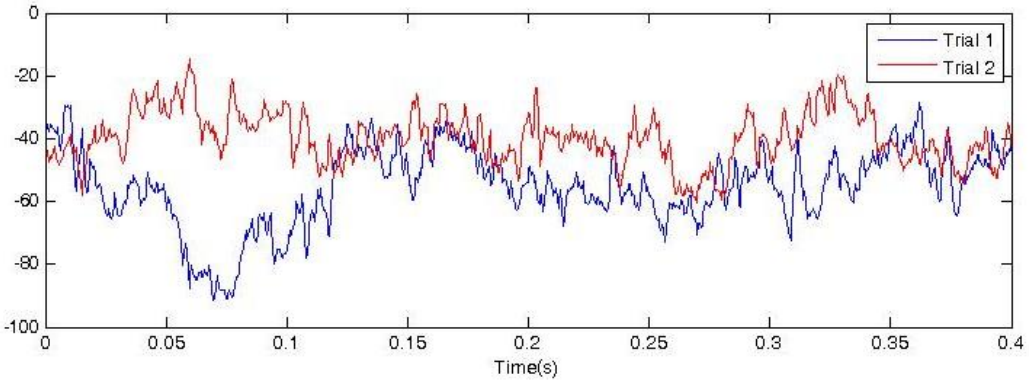
Inputs: EEG data (the `data` matrix), the time of each stimulus (the `events` matrix), an electrode number, and two inputs to define the time interval that you will extract before/after each stimulus (in seconds, e.g. t1=0.2 and t2=0.6 will extract 200ms before the trial to 600ms after the trials ends)
Outputs: a time vector (in seconds) so you can plot the data segments you are extracting, and a single matrix where each row is one trial and the columns represent points in time.

Assume the data has already been loaded into the workspace. Include your code from #3 in the function, so the extra zeros and DC offsets will be removed before the trials are extracted.

Note that the `events` matrix contains six columns, representing the year/month/day/hour/minute/second that the stimulus was presented. To determine the amount of time that has elapsed from the first stimulus to the nth stimulus, use the command

```
elapsed = etime(events(n,:), events(1,:));
```

Hint: If you plot the first two trials from channel 1 in "sub8_sess4_6" it should look like this:



Once you have written the function, use it to answer these questions:

a. For the Pz electrode (channel 13), create a plot showing all of the "target" trials in red and the non-target trials in blue, over the time range [-0.2, 0.6] seconds. Create a second plot showing the averages (ERPs) of the target trials and the non-target trials. What do you notice about the individual trials? Do you see any evidence of the P300 in this data?

b. Recall that the data was recorded on a "reference-free" system and that we must choose an appropriate reference to correctly interpret the data. Add code to your function to re-reference the data to the mean of channels T7 and T8 (indices 7 and 24).

   Now recreate the plots from (a) using the re-referenced data. You should be able to see evidence of the P300 in the ERPs, even though it is not visible on a trial-by-trial basis.

c. What could you do differently in order to produce cleaner and more convincing evidence of the P300?

5. Write a function called "`myTopoPlot`" to create a topographical plot of the data from all 32 electrodes at a single point in time:

```
myTopoPlot(EEG, chanLocs)
```

Inputs: EEG data (one data point for every electrode location) and channel locations (from the .mat file "channel_locations" which has been provided to you)
Outputs: none

Your function should follow this procedure:

- Convert the `chanLocs32.theta` and `chanLocs32.radius` variables to vectors using these two lines of code:

```
thetaVec = [chanLocs32.theta];
```

```
radiusVec = [chanLocs32.radius];
```

- The theta angle measurement is in degrees; convert it to radians.
- Use the function `pol2cart` (or the mathematical formulas) to convert theta and radius to "x" and "y" in Cartesian coordinates.
- Find the minimum and maximum values of x and y. Then use the `linspace` function to create one vector spanning x_{min}:x_{max} and one vector spanning y_{min}:y_{max}.
- Use these vectors as input to the `meshgrid` function. The outputs will be two matrices, one denoting the x-coordinate at all points of the grid, and the other denoting the y-coordinate at all points of the grid.
- Use the `griddata` function to calculate the interpolated EEG data. The inputs will be the Cartesian electrode locations, a vector of EEG data, and the matrices from the meshgrid function. The output will be the interpolated EEG data.
- Use the function `contourf` to plot the interpolated EEG data. *In order for your function to match topoplot, you will need to input the Y-axis first:* *`contourf(yRange, xRange, interpolatedEEG).`* Use the command `axis square` to ensure that the x- and y-axis have the same spacing.

Hint: Here, we are essentially recreating the `topoplot` function from EEGLab so you understand exactly what the function is doing. So if the output of `topoplot` approximately matches the output from your function, you'll know that yours is working correctly.

Once your function is working, answer the following questions:

a. Calculate the average responses (ERPs) for targets and non-targets for each electrode. This is exactly the same as part 4.a. except now you are doing the calculation for each channel. You should have 32 target ERPs and 32 non-target ERPs, one for each channel.

b. Use your `myTopoPlot` function to create topographical plots of the ERPs: Calculate the mean ERP value (for both targets and non-targets) over the time range [-0.2, 0] seconds prior to the stimulus, then plot the mean value for each channel location. This is a baseline measurement of the EEG activity. Is there a difference between the target and non-target responses leading up to the stimulus?

   Use the `colorbar` command to show the values associated with each color in the plots. If the limits are not the same, you may need to change them using the command `set(gca,'clim',[-x1 x2])` where x1 and x2 are the min and max color limits.

c. Now create topographical plots of the mean ERPs over the time range [0.2, 0.4] seconds following the stimulus. Do you see a difference between the target and non-target responses during this time interval?