BE 521: Homework 3 Questions

Feature extraction

Spring 2020

68 points

Due: Thursday, 2/13/2020 11:59pm

Objective: Extract features from data and build a simple detector

John Talley
Collaborators: Joseph Iwasyk

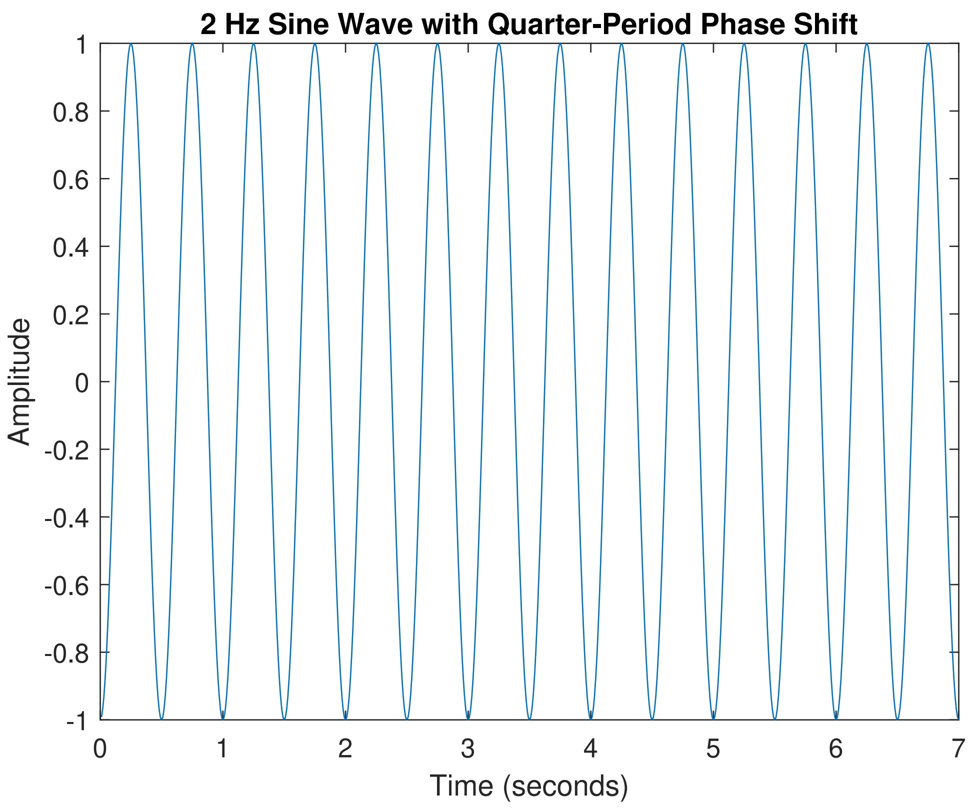
1 Features and Simulations (39 pts)

As you learned in class, features are the backbone of almost all detection strategies, from seizures in EEG to faces in images. Features are usually defined in journal articles as an equation or set of equations, and the task for the reader—if she wants to use that feature—is to implement that feature in code. In this section, you will explore and implement some features commonly used in EEG analysis and test them on simulated time-series data.

1. Consider the following toy signal: 7 seconds of a 2 Hz sine wave with a quarter period phase-shift, sampled at 100 Hz

   (a) Plot the signal. (2 pts)

```
fs = 100; % HZ
t = 1/fs:1/fs:7; % 100 Hz sampling rate
x = sin(4*pi*t - (pi/2));
figure;
plot(t, x);
xlabel('Time (seconds)');
ylabel('Amplitude');
title('2 Hz Sine Wave with Quarter-Period Phase Shift');
```



(b) Using the Matlab functions for the difference, sum, and absolute value of the elements in a vector (look them up if you don't know them), create an anonymous function for the line-length feature $LL(\mathbf{x}) = \sum_{i=2}^{n} |x_i - x_{i-1}|$ in one line of Matlab code that uses no loops (i.e., the outputs of one function will be the inputs of another). Your function should look something like

```
LLFn = @(x) XXXXXX;
```

where `XXXXX` represents some use of the aformentioned functions and the input signal `x`. (4 pts)

(c) What is the line length of this signal? (2 pts)

```
LLFn = @(x) sum(abs(diff(x)));
len = LLFn(x);
% length = 55.9921
```

Length = 55.9921

2. Consider line length of the signal using a sliding window with a certain amount of window overlap (or, to think of it another way, displacement with each "slide"). Now, instead of having just one value for the line length, you will have a number of values.

(a) Given a signal `x` with sampling frequency `fs` and windows of length `winLen` and displacement `winDisp` (both in seconds), create an anonymous function called `NumWins` that calculates the number of possible (full) windows in your signal of length `xLen` (in samples), i.e.,

```
NumWins = @(xLen, fs, winLen, winDisp) XXXXXX;
```

where XXXXXX is the single-line expression for this value. You may assume that **winDisp** is a factor of both **winLen** (as it usually is/should be) and the length (in seconds) of **x**. (4 pts)

```
NumWins = @(xLen, fs, winLen, winDisp) floor((xLen/fs - winLen)/winDisp + 1);
```

(b) Use this function to calculate the number of windows for the signal described in Question 1.1 for a 500 ms window with 250 ms displacement, i.e., the expression

```
      NumWins(length(x), fs, winLen, winDisp)
```

where **fs**, **winLen**, and **winDisp** are the appropriate values. (1 pt)

```
fs = 100; % Hz
winLen = 0.500; % seconds
winDisp = 0.250; % seconds

windows = NumWins(length(x), fs, winLen, winDisp); % 27 windows
```

27 Windows

(c) Repeat the above calculation for 50 ms window displacement. (1 pt)

```
winDisp = 0.050; % seconds
windows = NumWins(length(x), fs, winLen, winDisp); % 131 windows
```

131 Windows

(d) Repeat the above calculation for 100 ms window displacement. (1 pt)

```
winDisp = 0.100; % seconds
windows = NumWins(length(x), fs, winLen, winDisp); % 66 windows
```

66 Windows

3. (a) Create a function (in another file) called **MovingWinFeats(x, fs, winLen, winDisp, featFn)** that returns a vector of the values of the feature on the signal **x** in all the possible windows, where **featFn** is a feature function like the one you wrote in Question 1.1.b. You may find it useful to use your **NumWins** function (or at least its expression). You may assume that the product of **winDisp** and the sampling rate **fs** is an integer. (6 pts)

Make sure your MovingWinFeats code is in your pdf. One way is to use the following Matlab code (in your script) to automatically load in the function's code (where we assume that the function is one directory up from the *.tex file). Windows users may need to change the forward-slash to a backslash.

```
function features = MovingWinFeats(x, fs, winLen, winDisp, featFn)

num_windows = floor((length(x)/fs - winLen)/winDisp + 1);
x_start = 1;
winLen = fs*winLen;
winDisp = fs*winDisp;

for i = 1:num_windows
```

```
    x_end = winLen + (i-1)*winDisp;
    feature = x(x_start:x_end);
    features(i) = featFn(feature);
    x_start = x_end - (winLen - winDisp) + 1;
end

end
```

(b) Using the signal you defined in Question 1.1 and the function you created in Question 1.1.b, calculate the line-length over windows of length 500 ms and displacement 250 ms. (2 pts)

```
fs = 100; % Hz
winLen = 0.500; % seconds
winDisp = 0.250; % seconds
feats = MovingWinFeats(x, fs, winLen, winDisp, LLFn); % all 4.00 units long
totalLen_feats = sum(feats);
% total length of all features =  107.7871
```

Total length of all features = 107.7871

Lengths over 27 Windows: 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448, 3.99211470131448

(c) Add a unit-amplitude 5 Hz signal (in the form of a sine wave) to your original signal and again calculate the line length over the same window and displacement. (2 pts)

```
x = sin(4*pi*t - (pi/2)) + sin(10*pi*t); % Adding 5 Hz Unit Amplitude Sine Wave
features = MovingWinFeats(x, fs, winLen, winDisp, LLFn);
totalLen_features = sum(features);
% total length of all features = 275.2876
```

Total length of all features = 275.2876

Lengths over 27 Windows: 9.62615006927222, 10.3230954071745, 10.5156643211075, 10.3388660045455, 9.62615006927221, 10.3230954071745, 10.5156643211075, 10.3388660045455, 9.62615006927223, 10.3230954071745, 10.5156643211075, 10.3388660045455, 9.62615006927221, 10.3230954071745, 10.5156643211076, 10.3388660045455, 9.62615006927224, 10.3230954071745, 10.5156643211076, 10.3388660045455, 9.62615006927222, 10.3230954071744, 10.5156643211075, 10.3388660045455, 9.62615006927219, 10.3230954071744, 10.5156643211075

4. Code the following 3 additional features in MINIMAL lines of code (hint: each can be implemented in one line using the anonymous function trick).

(a) Area, $A(\mathbf{x}) = \sum_{i=1}^{n} |x_i|$ (2 pts)

(b) Energy, $E(\mathbf{x}) = \sum_{i=1}^{n} x_i^2$ (2 pts)

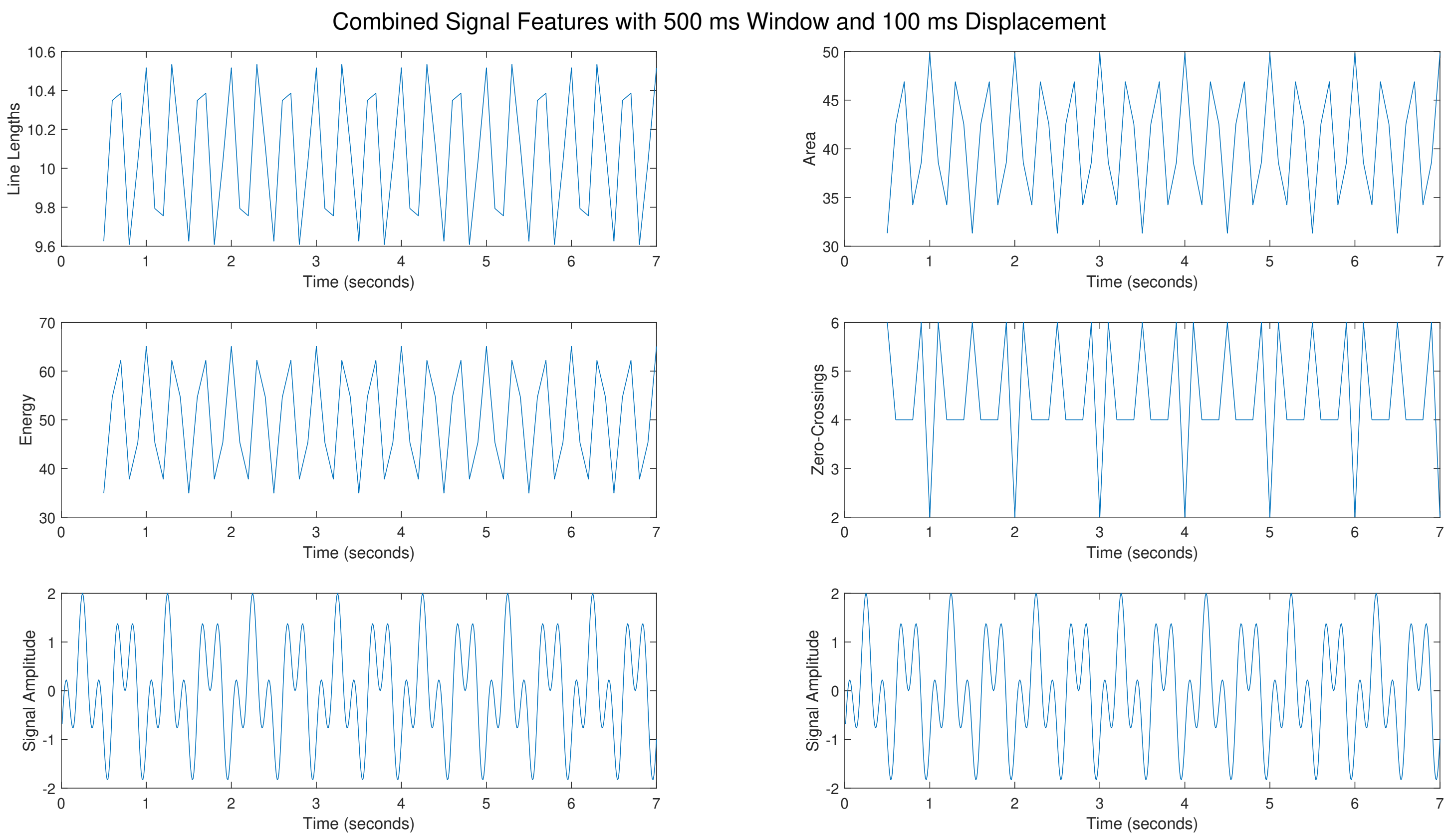
(c) Zero-Crossings around mean,

$ZX(\mathbf{x}) = \sum_{i=2}^{n} \mathbf{1}(\textbf{FromAbove})$ OR $\mathbf{1}(\textbf{FromBelow})$, where $\mathbf{1}(\cdot)$ denotes the indicator function, which returns a zero if its argument is false and a one if it is true, **FromAbove** denotes $(x_{i-1} - \overline{x} > 0)$ AND $(x_i - \overline{x} < 0)$, **FromBelow** denotes $(x_{i-1} - \overline{x} < 0)$ AND $(x_i - \overline{x} > 0)$, and $\overline{x}$ is the mean value of the elements in x. (4 pts)

```
Area = @(x) sum(abs(x));
area = Area(x);

Energy = @(x) sum(x.^2);
energy = Energy(x);

ZX = @(x) sum((x(1:length(x)-1) > mean(x) & x(2:length(x)) < mean(x)) | (x(1:length(x)-1) < mean(x) & x
zero_x = ZX(x);
```

(d) Plot the values of the four features on the combined signal in the first four cells of a 3x2 matlab subplot. Use a 500 ms window with 100 ms displacement. Using the right-aligned convention (where the "official" time of the feature is that of the last data point in the window), give the appropriate time axis for each window point. In addition, plot the original signal with the 2Hz and 5Hz components in the last two cells of the 3x2 subplot (to make comparing down the column easy). Ensure that the time axis in all of your plots is the same. (6 pts)

```
winLen = 0.500; % seconds
winDisp = 0.100; % seconds

feature_lengths = MovingWinFeats(x, fs, winLen, winDisp, LLFn);
feature_area = MovingWinFeats(x, fs, winLen, winDisp, Area);
feature_energy = MovingWinFeats(x, fs, winLen, winDisp, Energy);
feature_zx = MovingWinFeats(x, fs, winLen, winDisp, ZX);

tspan = 0.5:0.1:7;
figure('units','normalized','outerposition',[0 0 1 1]);
subplot(3,2,1);
plot(tspan, feature_lengths);
xlabel('Time (seconds)');
ylabel('Line Lengths');
subplot(3,2,2);
plot(tspan, feature_area);
xlabel('Time (seconds)');
ylabel('Area');
subplot(3,2,3);
plot(tspan, feature_energy);
xlabel('Time (seconds)');
ylabel('Energy');
subplot(3,2,4);
plot(tspan, feature_zx);
xlabel('Time (seconds)');
ylabel('Zero-Crossings');
subplot(3,2,5);
plot(t, x);
xlabel('Time (seconds)');
ylabel('Signal Amplitude');
subplot(3,2,6);
plot(t, x);
xlabel('Time (seconds)');
ylabel('Signal Amplitude');
sgtitle('Combined Signal Features with 500 ms Window and 100 ms Displacement');
```



2 Feature Overlays (17 pts)

In this section, you will use a line-length feature overlay on a segment of EEG containing a seizure. This data is stored in I521_A0003_D001

1. What is the length using hours:minutes:seconds:milliseconds of the recording? (Use getDuration) (2 pts)

```
dataset_ID = 'I521_A0003_D001';
ieeg_id = 'jtalley';
ieeg_pw = 'jta_ieeglogin.bin';

session = IEEGSession(dataset_ID,ieeg_id,ieeg_pw);

Fs  = session.data.sampleRate;
durationInUSec = session.data.rawChannels.get_tsdetails.getDuration;
durationInSec = durationInUSec/1e6;
duration = seconds(durationInSec);
duration.Format = 'hh:mm:ss.SSS';
% Duration = 01:21:20.390
```

Duration: 01:21:20.390

```
Warning: Objects of edu/upenn/cis/db/mefview/services/TimeSeriesDetails class
exist — not clearing java
Warning: Objects of edu/upenn/cis/db/mefview/services/TimeSeriesInterface class
```

```
exist — not clearing java
IEEGSETUP: Found log4j on Java classpath.
URL: https://www.ieeg.org/services
Client user: jtalley
Client password: ****
```

2. How many data points should we discard at the end if we want to clip the recording to the last full second? Do this clipping. (1 pt)

```
clip = Fs*0.390; % seconds —> samples
% 78 datapoints
```

Discard 78 data points

3. If we want to overlay a feature trace on the original signal, we have to interpolate that feature (which has been calculated over windows) for each data point of the original signal. One of the simplest methods of doing this is called zero-order interpolation, where we just hold the value constant until we get to the next calculated value. For example, if we had a moving window of 1 second with 1 second displacement, the zero-order interpolated feature vector would have the same value the entire first second, then the same for the entire second second, etc, where each second contains the same number of points as the sampling frequency of the original signal.

   (a) Using the `repmat` and `reshape` functions, create an external function `zoInterp(x, numInterp` that copies each value of `x` `numInterp` times. You can implement this function in one line of code with no loops. Include the code for this function as you did in Question 1.3.a. (2 pts)

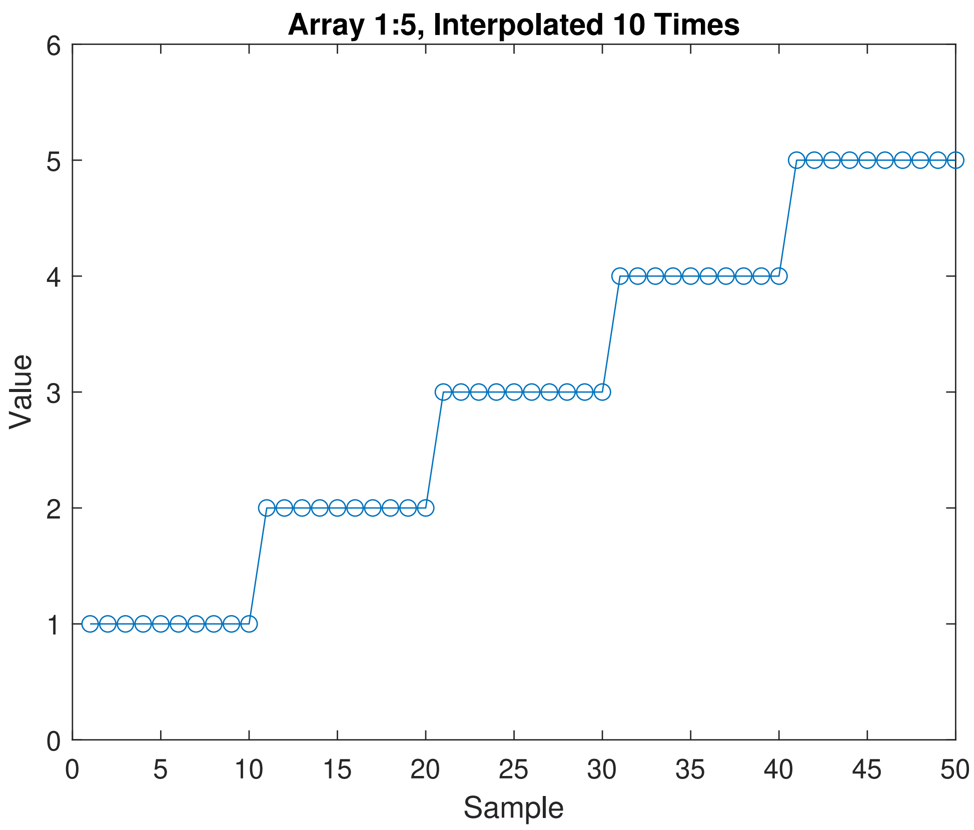
```
function x_interp = zoInterp(x, numInterp)
    x_interp = reshape(repmat(x,numInterp,1),1,length(x)*numInterp);
end
```

   (b) Confirm that this function works correctly by expanding the length of the vector `1:5` by a factor of 10 and plotting with the command

```
    plot(zoInterp(1:5,10),'—o')
```

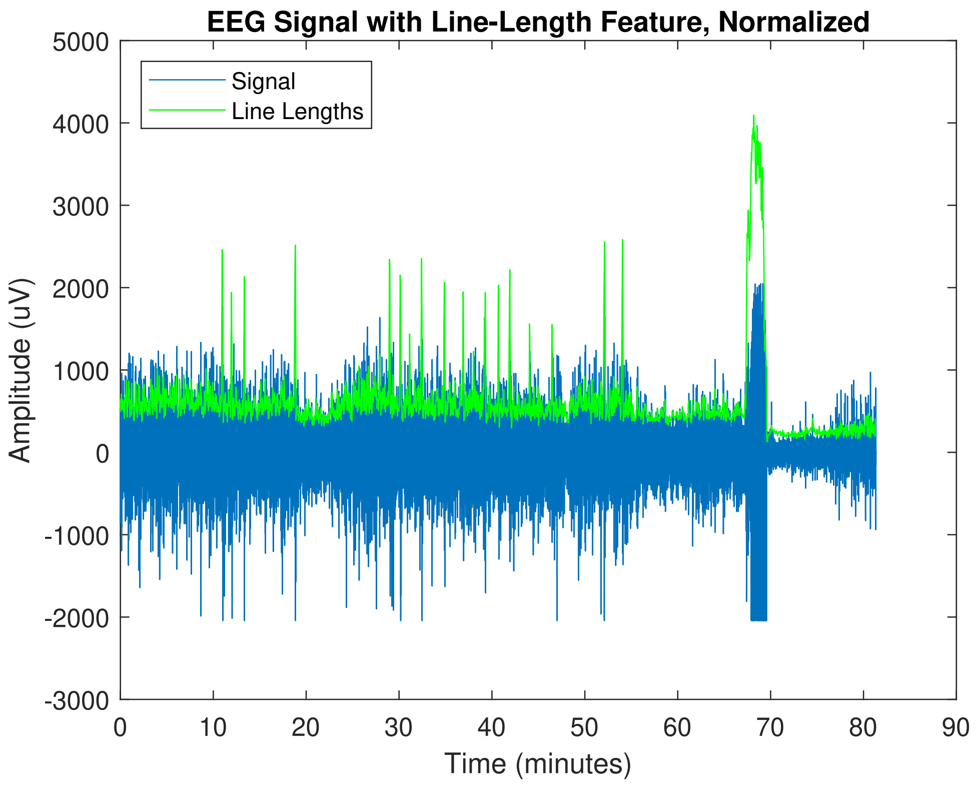
   where the `'-o'` option lets us see the individul points as well as the line that connects them. (2 pts)

```
figure;
plot(zoInterp(1:5,10),'—o');
ylim([0 6]);
xlabel('Sample');
ylabel('Value');
title('Array 1:5, Interpolated 10 Times');
```



4. Using a 4-second sliding window with 1-second displacement, calculate the line length feature over the entire signal. Normalize the line-length feature values to have a maximum twice that of the original EEG signal maximum. Plot the signal in blue and overlay the right-aligned line-length feature in green. Note: you will need to pad your signal in order to get them to line up correctly and be the same length. Put the units of your time axis in minutes, and be sure to add a legend in a location in the plot that does not cover up any signal or feature. (6 pts)

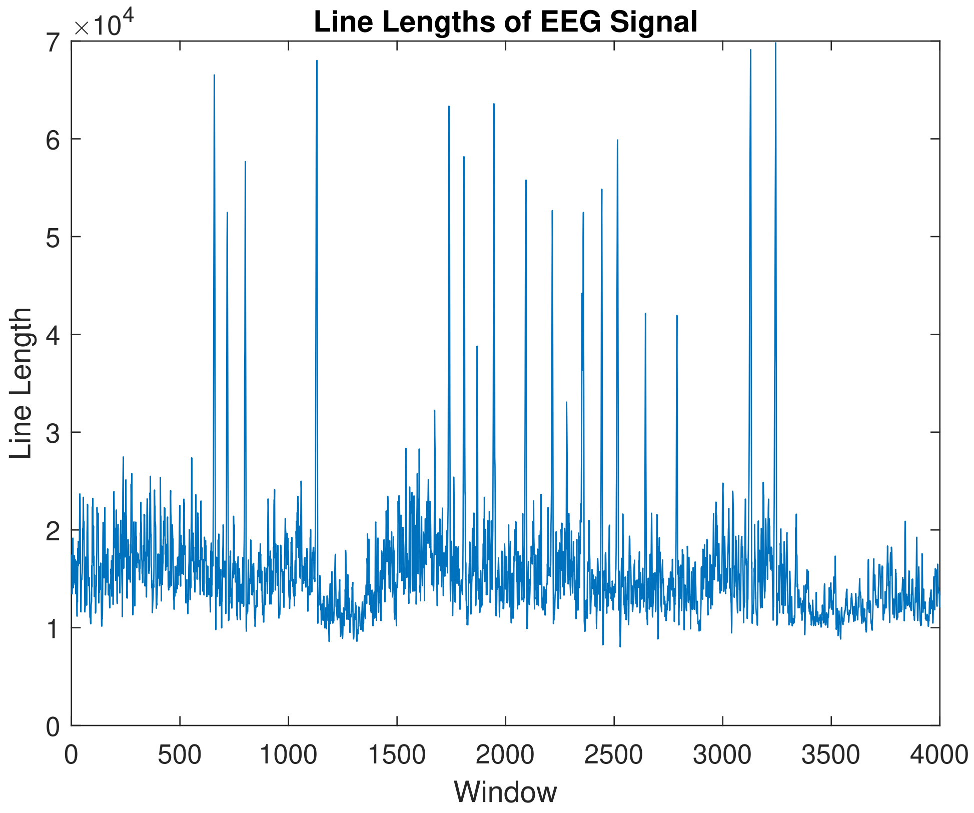
```
signal = getvalues(session.data,1:(durationInSec*Fs-clip),1);
time = 1/(Fs*60):1/(Fs*60):(durationInSec/60 - clip/(Fs*60));

winLen = 4.0; % seconds
winDisp = 1.0; % seconds
signal_line_lengths = MovingWinFeats(signal, Fs, winLen, winDisp, LLFn);
timespan = 4.0:1.0:durationInSec;

signal_line_lengths_pad = zoInterp(signal_line_lengths, Fs);
time_span = 4.0/(Fs*60):1.0/(Fs*60):((length(signal_line_lengths_pad)+3)/(Fs*60));

scalar = 2*(max(signal)/max(signal_line_lengths));
signal_line_lengths_norm = scalar.*(signal_line_lengths_pad);

figure;
plot(time, signal);
hold on;
plot(time_span, signal_line_lengths_norm,'g');
xlabel('Time (minutes)');
ylabel('Amplitude (uV)');
title('EEG Signal with Line-Length Feature, Normalized');
legend('Signal', 'Line Lengths','Location','northwest');
```



5. What threshold might you use on the raw line-length feature vector (not the normalized one used for plotting) in order to capture the 17 largest pre-seizure chirps that occur? (1 pt)

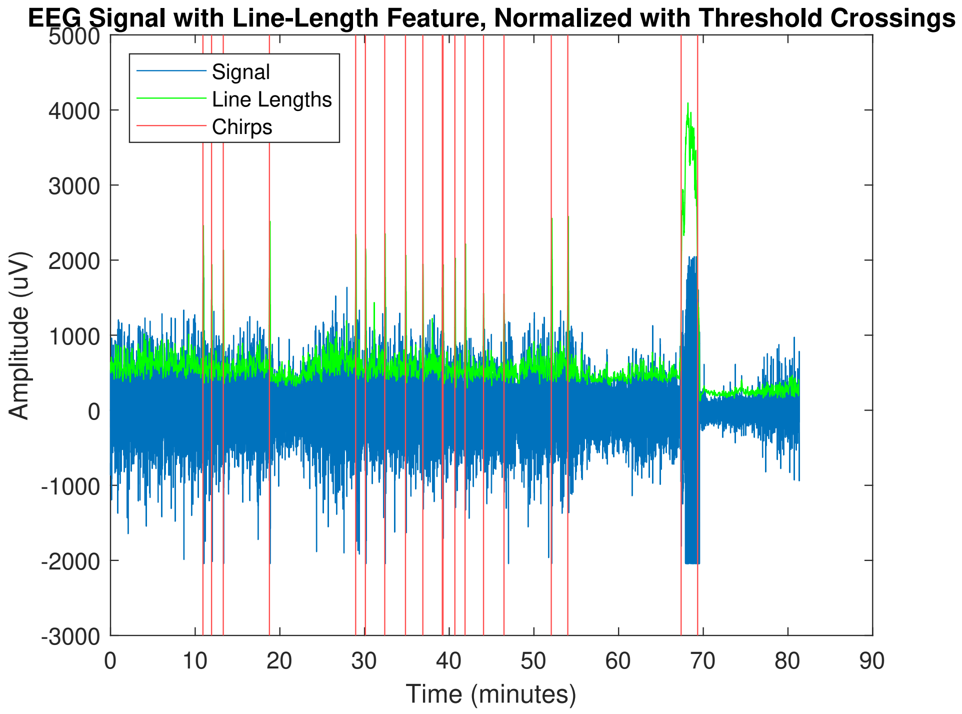
```
% Visualize the raw line—length feature pre—seizure to determine threshold
figure;
plot(signal_line_lengths(1:4000));
xlabel('Window');
ylabel('Line Length');
title('Line Lengths of EEG Signal');
% 4.1e4 seems like a reasonable threshold
% Use findpeaks to verify threshold (excluding seizure event)
[~,locs] = findpeaks(signal_line_lengths(1:4000), 'MinPeakHeight', 4.1e4);
% Confirmed 17 peaks with a cutoff threshold of 4.1e4
```

Threshold: 41,000 uV



6. Using this threshold value, in another plot draw red vertical lines at the leading point in time where the threshold is crossed. Add these vertical lines on top of the plot you made in Question 2.4. These events should capture the pre-seizure chirps, the seizure onset, and some flickering during the end of the seizure. (3 pts)

```
% Use Zero-Crossings, but replace mean with threshold

ThreshX = @(signal_line_lengths_pad) find((signal_line_lengths_pad(1:length(signal_line_lengths_pad)-1) < 4.
thresh_x = ThreshX(signal_line_lengths_pad);
% convert threshold crossing points to minutes
thresh_x = thresh_x./(Fs*60);

figure;
plot(time, signal);
hold on;
plot(time_span, signal_line_lengths_norm,'g');
hold on;
for i = 1:length(thresh_x)
    xline(thresh_x(i),'r');
end
xlabel('Time (minutes)');
ylabel('Amplitude (uV)');
title('EEG Signal with Line-Length Feature, Normalized with Threshold Crossings');
legend('Signal', 'Line Lengths', 'Chirps','Location','northwest');
```



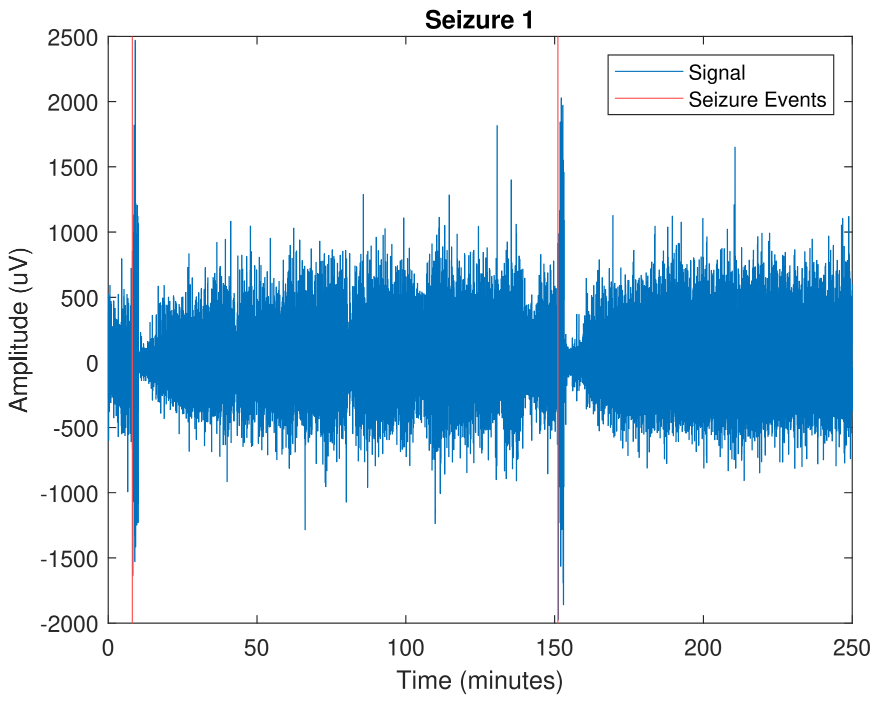
3 Building a Detector (12 pts)

In this section, you will use the features you defined previously to build a seizure detector. Use the EEG data in the file `I521_A0003_D002` with channels `multiSz_1`, and `multiSz_2`.

1. Plot the signal in `multiSz_1` and draw vertical red lines at the times when you think the two seizures begin. (You should be able to do this without the need of any features.) (2 pts)

```
dataset_ID = 'I521_A0003_D002';
session2 = IEEGSession(dataset_ID,ieeg_id,ieeg_pw);

durInUSec1 = session2.data.rawChannels(1).get_tsdetails.getDuration;
durInSec1 = durInUSec1/1e6;
durInUSec2 = session2.data.rawChannels(2).get_tsdetails.getDuration;
durInSec2 = durInUSec2/1e6;

Fs2  = session2.data.sampleRate;

Sz1 = getvalues(session2.data,1:durInSec1*Fs2,1);
Sz2 = getvalues(session2.data,1:durInSec2*Fs2,2);

timeSpan = 1/(Fs2*60):1/(Fs2*60):durInSec1/60;
figure;
plot(timeSpan, Sz1);
hold on;
xline(8.1,'r'); % seizure 1, eyeballed from graph
xline(151.1,'r'); % seizure 2, eyeballed from graph
xlabel('Time (minutes)');
ylabel('Amplitude (uV)');
title('Seizure 1');
legend('Signal', 'Seiure Events');
```

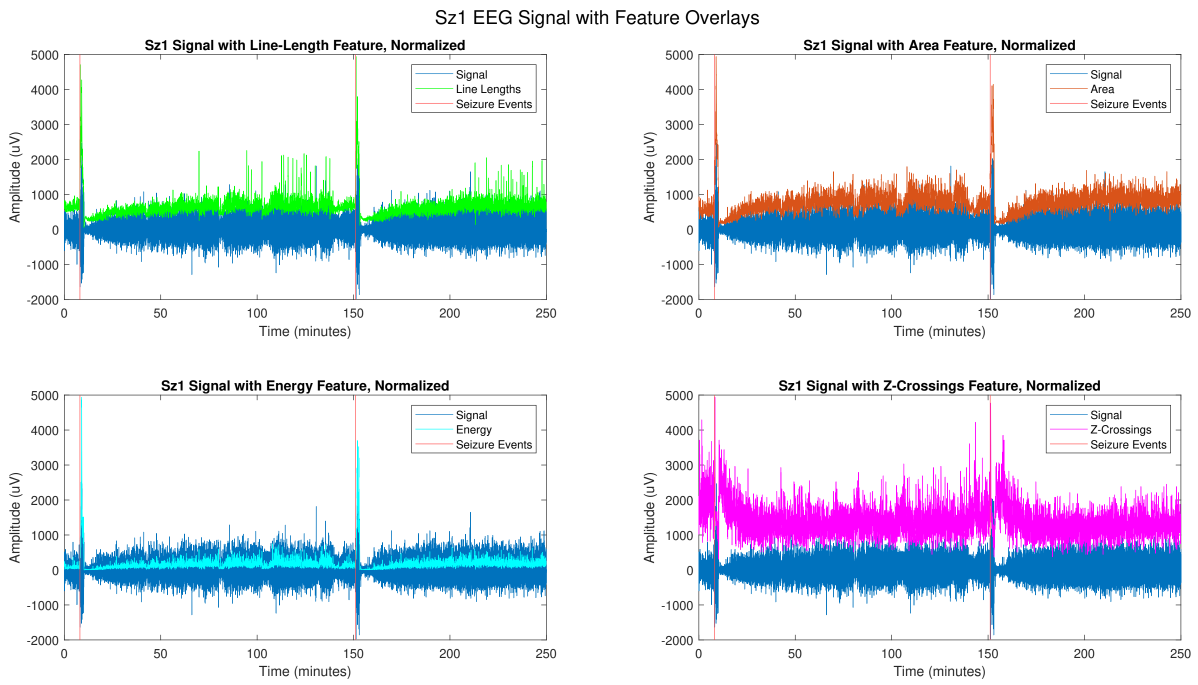
```
Warning: Objects of edu/upenn/cis/db/mefview/services/TimeSeriesDetails class
exist — not clearing java
Warning: Objects of edu/upenn/cis/db/mefview/services/TimeSeriesInterface class
exist — not clearing java
IEEGSETUP: Found log4j on Java classpath.
URL: https://www.ieeg.org/services
Client user: jtalley
Client password: ****
```



2. Produce feature overlay plots similar to that of Question 2.4 for each of the four features you have implemented along with the red vertical lines at each seizure. Use the same 4-second sliding window with 1 second displacement. (4 pts)

```
Sz1_line_lengths = MovingWinFeats(Sz1, Fs2, winLen, winDisp, LLFn);
Sz1_area = MovingWinFeats(Sz1, Fs2, winLen, winDisp, Area);
Sz1_energy = MovingWinFeats(Sz1, Fs2, winLen, winDisp, Energy);
Sz1_zx = MovingWinFeats(Sz1, Fs2, winLen, winDisp, ZX);

Sz1_line_lengths_pad = zoInterp(Sz1_line_lengths, Fs2);
Sz1_area_pad = zoInterp(Sz1_area, Fs2);
Sz1_energy_pad = zoInterp(Sz1_energy, Fs2);
Sz1_zx_pad = zoInterp(Sz1_zx, Fs2);
time_Span = 4.0/(Fs2*60):1.0/(Fs2*60):((length(Sz1_line_lengths_pad)+3)/(Fs2*60));

scalar = 2*(max(Sz1)/max(Sz1_line_lengths));
Sz1_line_lengths_norm = scalar.*(Sz1_line_lengths_pad);

scalar = 2*(max(Sz1)/max(Sz1_area));
Sz1_area_norm = scalar.*(Sz1_area_pad);

scalar = 2*(max(Sz1)/max(Sz1_energy));
Sz1_energy_norm = scalar.*(Sz1_energy_pad);
```

```
scalar = 2*(max(Sz1)/max(Sz1_zx));
Sz1_zx_norm = scalar.*(Sz1_zx_pad);

figure('units','normalized','outerposition',[0 0 1 1]);
% Line Lengths
subplot(2,2,1);
plot(timeSpan, Sz1);
hold on;
plot(time_Span, Sz1_line_lengths_norm,'g');
hold on;
xline(8.1,'r'); % seizure 1, eyeballed from graph
xline(151.1,'r'); % seizure 2, eyeballed from graph
xlabel('Time (minutes)');
ylabel('Amplitude (uV)');
title('Sz1 Signal with Line-Length Feature, Normalized');
legend('Signal', 'Line Lengths', 'Seizure Events','Location','northeast');
% Area
subplot(2,2,2);
plot(timeSpan, Sz1);
hold on;
plot(time_Span, Sz1_area_norm);
hold on;
xline(8.1,'r'); % seizure 1, eyeballed from graph
xline(151.1,'r'); % seizure 2, eyeballed from graph
xlabel('Time (minutes)');
ylabel('Area (uV)');
title('Sz1 Signal with Area Feature, Normalized');
legend('Signal', 'Area', 'Seizure Events','Location','northeast');
% Energy
subplot(2,2,3);
plot(timeSpan, Sz1);
hold on;
plot(time_Span, Sz1_energy_norm,'c');
hold on;
xline(8.1,'r'); % seizure 1, eyeballed from graph
xline(151.1,'r'); % seizure 2, eyeballed from graph
xlabel('Time (minutes)');
ylabel('Energy (uV^2*sec)');
title('Sz1 Signal with Energy Feature, Normalized');
legend('Signal', 'Energy', 'Seizure Events','Location','northeast');
% Z-crossings
subplot(2,2,4);
plot(timeSpan, Sz1);
hold on;
plot(time_Span, Sz1_zx_norm,'m');
hold on;
xline(8.1,'r'); % seizure 1, eyeballed from graph
xline(151.1,'r'); % seizure 2, eyeballed from graph
xlabel('Time (minutes)');
ylabel('Z-Crossings');
title('Sz1 Signal with Z-Crossings Feature, Normalized');
legend('Signal', 'Z-Crossings', 'Seizure Events','Location','northeast');
sgtitle('Sz1 EEG Signal with Feature Overlays');
```



3. (a) Based on your plots in the previous question, which of the four features seems to give the largest signal (relative to the background) for when a seizure occurs? Explain why you think this feature is the best. (3 pts)

   Relative to the background, the Energy feature gives the largest signal at the occurence of each seizure event. This feature also serves as a good indicator for seizure detection because it appears to have the strongest signal to noise ratio (SNR) at each seizure event, which would help to minimize the false positive / missed detection rate. This will increase the accuracy of our detector and yield a more sensitive ROC curve.

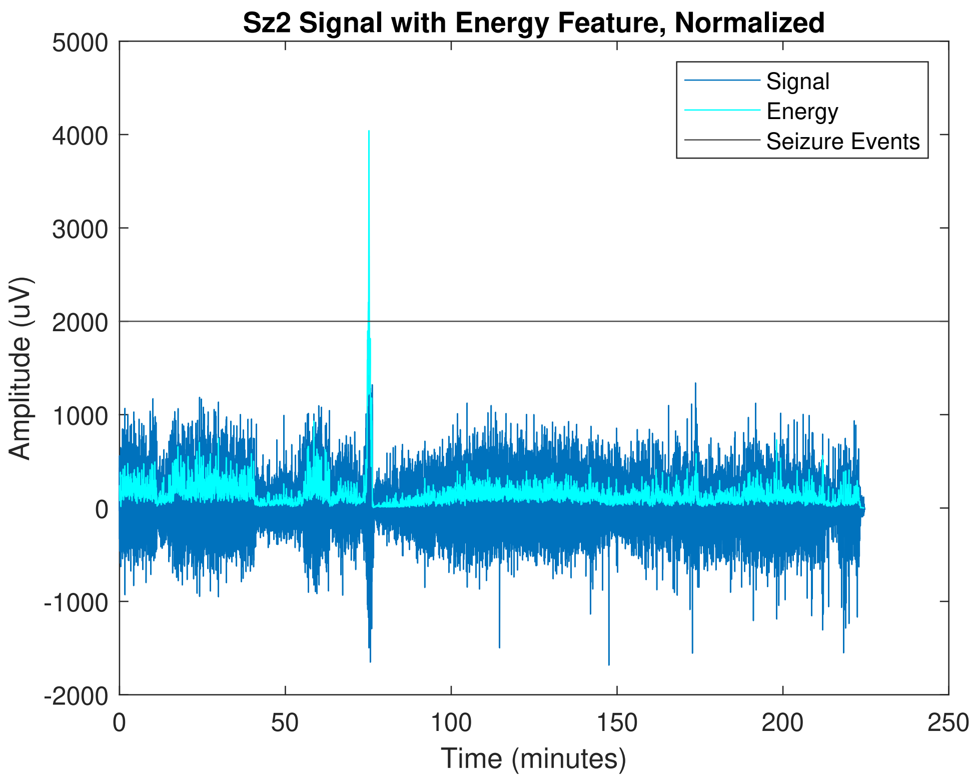
   (b) What threshold would you use to determine if a seizure is occurring? (1 pt)

   Using the Energy feature, 2000 uV is a reasonable threshold for seizure detection.

4. The signal in `multiSz_2` contains another seizure (whose location should again be fairly obvious). Plot the data along with the feature and threshold (horizontal black line, with correct normalization for the signal in `data2`) you determined in the previous question. (2 pts)

```
timeSpan2 = 1/(Fs2*60):1/(Fs2*60):durInSec2/60;
time_Span2 = 4.0/(Fs2*60):1.0/(Fs2*60):((length(Sz2_energy_pad)+3)/(Fs2*60));

Sz2_energy = MovingWinFeats(Sz2, Fs2, winLen, winDisp, Energy);
Sz2_energy_pad = zoInterp(Sz2_energy, Fs2);
scalar = 2*(max(Sz2)/max(Sz2_energy));
Sz2_energy_norm = scalar.*(Sz2_energy_pad);

figure;
plot(timeSpan2, Sz2);
hold on;
plot(time_Span2, Sz2_energy_norm,'c');
yline(2000,'k');
xlabel('Time (minutes)');
ylabel('Amplitude (uV)');
```

```
title('Sz2 Signal with Energy Feature, Normalized');
legend('Signal', 'Energy', 'Seizure Events','Location','northeast');
% Yes, a seizure occurs
```



Yes, a seizure occurs.