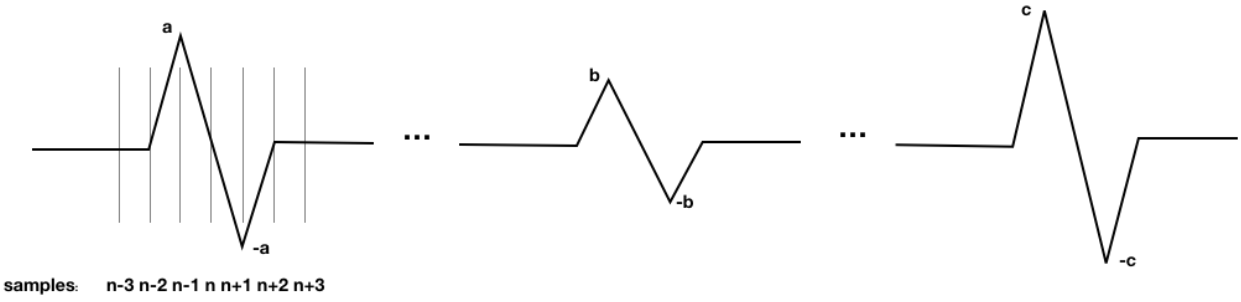
COREMAP DATA SCIENTIST CANDIDATE HOMEWORK PROBLEM

PROBLEM OVERVIEW

Event detection is a common task in biological signal processing. This project focuses on detecting specific events in the presence of noise and then characterizing event patterns.

The data set for this problem is a collection of waveforms that were simultaneously collected with a 12 x 9 grid of sensors. Each waveform is about 1 second long and was sampled at 1 kHz. The following diagram illustrates what one waveform would look like in the absence of noise:



Each event is composed of five samples that transition from zero to amplitude A, to zero, to amplitude -A, and back to zero. A waveform contains multiple separate events, each with a possibly different value of A. The time of an event is considered to be the central zero, labeled sample n in the example above. Waveform deflections of this form are considered physiological. All other features of the waveforms are considered non-physiological. During signal acquisition, the waveforms were corrupted by various types of noise.

PART I: EVENT DETECTION

The first part of the problem is to automatically detect all physiological events in the provided data file. The data format is CSV, with the first column being sample times and the rest of the columns being waveforms from each sensor. The first row contains sensor labels, which are of the form “s#r,#c” where #r is the row number (0-11) and #c is the column number (0-8).

Write an algorithm that detects the physiological events described above within each waveform. For each detected event, report the event time along with a quantitative value or qualitative label that captures the degree of confidence that the corresponding event is genuine. This assessment of confidence should consider the entirety of the data set, not just the individual feature (see part III below). Report these times and values / labels in a structured text file (such as JSON or clearly labeled plain text).

Include the following in the project report or presentation:

1) Summary of method of detection
2) Description of key parameters and their effect on detection accuracy
3) Overall, how accurate is the detection system?

PART II: WAVEFORM DESCRIPTION

As part of the project report or presentation, describe the types of noise which are present in the waveforms. Noise is considered any deviation from the ideal waveform illustrated above. For each type of noise, provide the following:

1) Explain the type of noise, including any important characteristics or parameters
2) Was it necessary to attenuate this noise to improve event detection? If so, how was the noise reduced or eliminated? How effective was this method?
3) What effect did this type of noise (after any attenuation) have on event detection accuracy?

PART III: EVENT PATTERNS

For this part of the problem, characterize any patterns or relationships between the individual events when the dataset is considered as a whole. Report the following:

1) Any observed patterns or relationships in the events
2) Whether these observations alter the degree of confidence that some events are genuine
3) Do these observations suggest ways to improve detection accuracy? If so, what specific methods could be used to realize these improvements?

PART IV: MACHINE LEARNING

Finally, assume physiological events have been correctly annotated in a large database of datasets such as this example. Describe how machine learning techniques could be applied to this event detection problem. Specifically:

1) What features would be used? How would these be extracted from the waveforms?
2) What machine learning methods would be useful? What outputs would these methods generate?
3) What benefits would these methods provide over classical signal processing techniques?

DELIVERABLES

Please submit the following:

1) A file containing event times along with a quantitative or qualitative rating of confidence that the event is genuine for each sensor
2) A report or presentation addressing the specific questions above
3) All analysis code used to complete this project