Homework 2 **CSCE 689**
Due: 11.59pm on October 1st, 2018
Total: 10pts

Problem 1: Normal and abnormal ECG signals

In this problem we will study selected electrocardiogram (ECG) signals from the MIT-BIH Arrhythmia Database, which contains data from healthy people and patients with cardiac diseases. The provided data contain normal and ventricular premature beats. The latter are caused when a heartbeat is not initiated by the sinuatrial node, the normal heartbeat initiator, but rather than different part of the heart, and tend to produce a skipped beat or odd chest sensation. You can find more details regarding the dataset and the ventricular premature beats in the following links: `https://en.wikipedia.org/wiki/Premature_ventricular_contraction` and `https://www.physionet.org/physiobank/database/mitdb/`.

The data that will be used for this homework can be found in the Google Drive under *CSCE689_Fall2018_ClassMaterial* $\rightarrow$ *Homework2* $\rightarrow$ *MitArrhythmia*. Folder *Data* contains the ECG values sampled at a sampling frequency of $F_s = 360Hz$. Folder *Annotations_QRSLocation* contains the location of QRS complexes for each signal denoted in samples. Folder *Annotations_NormalAbnormalBeats* includes labels of normal (N) and ventricular premature (V) beats (3rd column) and their corresponding times in seconds and samples (1st and 2nd column, respectively).

(a) (1 pt) Visualization of ECG components: Select and plot examples of ECG signals containing normal beats with duration of 10-20sec. Please note the three main ECG components, i.e., P wave, QRS complex, and T wave.

(b) (1 pt) Visualization of premature ventricular beats: Select and plot examples of ventricular beats with duration of 10-20sec.

(c) (2 pts) Pan-Tomkins QRS detection: Implement the Pan-Tomkins algorithm to detect the QRS complexes of all signals, i.e., bandpass filtering, differentiation, squaring, moving average filtering, peak detection. Using the annotation files in folder *Annotations_QRSLocation*, please provide the precision and recall measures between the detected and actual QRS complexes using a forgiving threshold of 0.5sec, i.e., a detected QRS complex is correct if it is at most 0.5sec distance apart from an actual QRS complex. Visualize your results for different ECG segments containing normal and premature ventricular beats and provide your observations.

(d) (2 pts) ECG feature extraction: Segment the ECG signals into 5sec analysis frames. Using the annotation files in folder *Annotations_NormalAbnormalBeats*, assign a normal/abnormal label to each 5sec analysis frame using a majority voting strategy, i.e., if the current frame contains more normal beats compared to ventricular premature beats, then it is considered normal, otherwise it is considered abnormal. You can ignore a 5sec segment if no annotations are available. For each analysis frame, please compute the number of beats, the interbeat interval, and the coefficients of a 8-point Fourier transform. Please compute the histograms of these features separately for the normal and abnormal 5sec frames.

(e) (2 pts) Bonus: Perform a two-sample t-test to find if there are significant differences between the normal and abnormal 5sec frames with respect to the ECG features computed

in question 1d. You can find more information regarding the t-test in this link: `https://towardsdatascience.com/inferential-statistics-series-t-test-using-numpy-2718f8f9bf2f`. *Tip:* Use the stats.ttest_ind function in scipy.

Problem 2: EDA reactivity to multimedia content

In this problem, we will examine individuals' reactions to multimedia content. Participants were instructed to watch several types of videos, while their physiological responses were being recorded. At the end of each video, they were also asked to provide their subjective ratings of valence and arousal on a continuous 9-point scale. You can find more information regarding this dataset in this link: `http://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html`.

The data for this problem is located in the Google Drive under *CSCE689_ Fall2018_ClassMaterial* $\rightarrow$ *Homework2* $\rightarrow$ *DEAP*. In folder *Data*, you can find the EDA values sampled at $F_s = 32Hz$ for each participant over 40 trials, e.g., file *s01_Trial10.csv* includes the EDA data from participant 01 during trial 10. The corresponding arousal/valence annotations can be found in file *participant_ratings.csv*, which include the participant ID (1st column), the number of trial (2nd column), and the self-reported valence and arousal annotations (3rd, 4th column, respectively).

(a) (1 pt) Visualization of EDA signals: Plot EDA signals from different participants and mark examples of the skin conductance responses (SCR). What do you observe? Is there variability in the data?

(b) (2 pts) Feature extraction: Extract the mean level (also called skin conductance level, SCL), SCR frequency, and mean SCR amplitude (i.e., the average amplitude of all SCRs) for each signal. Provide several examples of EDA signals and the resulting features.

(c) (1 pt) Relation between EDA features and self-reported ratings: Compute the correlation values between the extracted EDA features and self-reported arousal and valence ratings. Are the correlations significant? Please discuss your results and how you would improve it.