The ECG Challenge

Alastair, Timi, Weiyun and Ying

The University of Nottingham

Topics covered in this presentation

Introduction

Background Information

- ECG stands for an electrocardiogram.
- An ECG looks at a persons heart's rate, rhythm and electrical activity.
- When would you get an ECG?
  - If you have symptoms of a heart attack or coronary heart disease.
  - If you have a pre-existing heart condition.
  - While taking certain medication.
- Willem Einthoven is credited with inventing the ECG

Figure 1: An example of a graph from an ECG reading.

Overview of the Problem

Aim For This Challenge

Our Aim

- Create a model to diagnose two heart conditions:
    1. Myocardial infarction (MI), commonly known as a heart attack.
    2. Cardiomyopathy, the general term for diseases of the heart muscle.

Data Given To Us

- Training data:
    - ECGs for 115 patients.
    - Heart condition (Healthy, MI, Cardiomyopathy).
- Test data:
    - Just the ECGs for 100 patients.
- Our model should classify the training data correctly.

Example for a patient

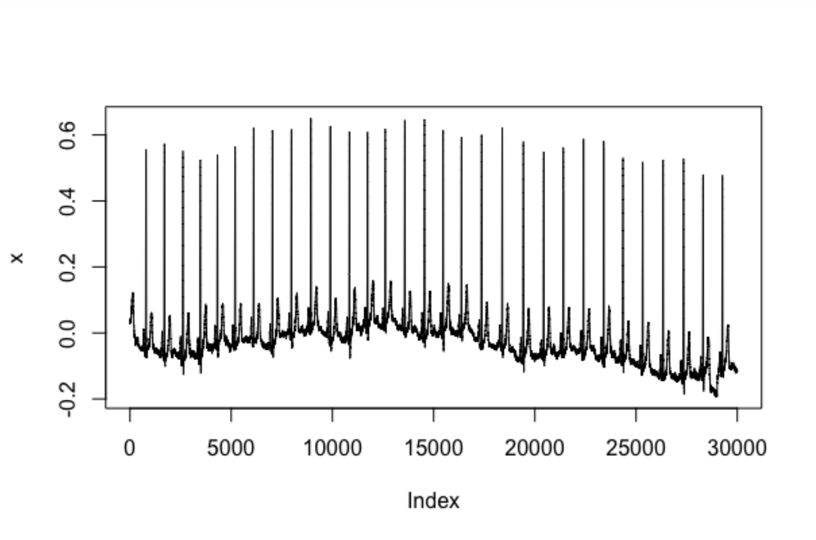


Figure 2: Plot of patient 1 from training data.

Challenges Faced So Far

Main Issue

Extracting features
What indicates issues with the heart?

- T-wave inversion, the t-wave goes down instead of up.
- QRS interval is lengthened.
- ST elevation, the heartbeat doesn't return to baseline quickly after QRS complex.

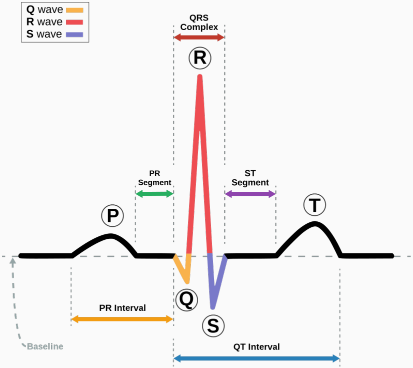


Figure 3: Diagram of a single heartbeat

Manipulating the Data

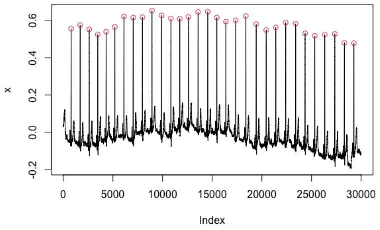
Splitting Into Individual Heartbeats

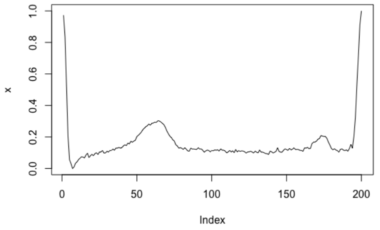
Determining maximums

- Set a threshold that the point must be above.
- Look for a point higher than the points either side of it.

Splitting at the maximums

- Split vector at the maximums
- Interpolate each heartbeat to be of length 200
- Scale y-axis to be in $[0, 1]$



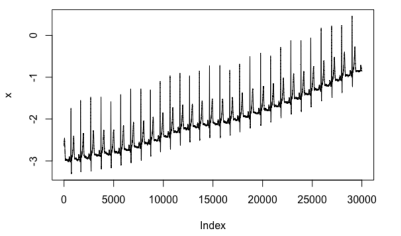


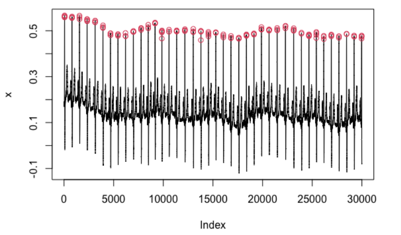
Drift in ECG machine

- Model drift with lm function in r.
- Use coefficients $c =$ y-intercept and $m =$ gradient to remove drift.
- $x \leftarrow x - m * i - c$

Multiple indices for each peak

- Create function to check the if index points are close together.
- Tolerance used was 100.

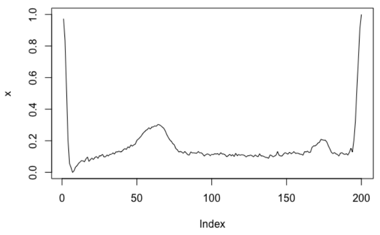




Smoothing the Functions

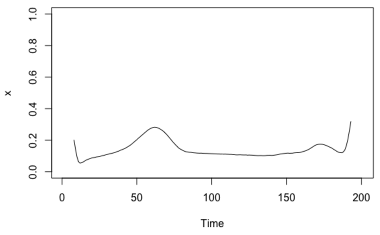
Method for smoothing function

- Use a moving averages method.
- Each new point is the average of the 15 data closest points to it.



Issues with this method

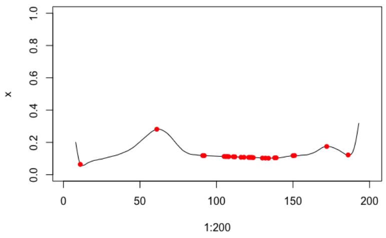
- First and last 7 points don't get values.
- Trade off when selecting window size.
    - Higher window size gives a smoother curve.
    - Higher window size can lead to loss of features in the graph.



Why Has All This Been Useful?

Using the derivative

- Now we have smoothed the function we can take the derivative.
- Plotted are the points where the derivative is equal to zero
- These minima and maxima can be used to extract features.



T-wave inversion

Look at the second derivative at the second turning point.

QRS interval

Look at the length of the QRS interval.

QRS Interval = position of first TP + (200 − position of last TP)

Predictions So Far

Using Classification Trees

Baseline predictions

- Baseline prediction from looking at the original data without manipulation.
- Look at the variance of the data for each patient.
- Gave us a prediction accuracy of 65% on the test set.

Classification Trees

- As with many applications in medicine this problem lends itself well to decision trees.
- This allow factors that are definite indicators of a certain condition to be accounted for.
- Random Forests can be used to improve accuracy of predictions.
- This gave us a prediction accuracy of 67%.

Conclusions and Further Work

Work still to do

Conclusions

- The key part of this challenge has been extracting features from the data.
- Decision trees and thus Random Forests seem to be a good model to use for predicting heart conditions.

Work still to do

- Continue looking for features to extract using exploratory plots and ways to extract them automatically.
- Explore other machine learning methods, potentially multinomial logistic regression?
- Use cross validation to compare these methods.

Thank You For Listening

Any Questions?