# Challenge:

To build a regression model that predicts the left ventricular heart volume (a scalar value, Y) from a large number of impedance measurements (a 6720-dimensional vector, X).

# Description of Data:

## Training data:

There are 77000 training samples.

***train***

The file ***train*** contains a 77000x6721 matrix, in CSV format. For each row, the first 6720 columns are the 6720 dimensions of the input vector, X. The last column is the output scalar, Y. Please see *Features* section below for an explantion of the 6720-dimension feature.

***trainOccl***

The file ***trainOccl***contains a 77000x32 matrix, in CSV format. Each row provides information on which electrodes (if any) were occluded for the corresponding ***train*** measurements. For example, the 1st row of ***trainOccl***is

[0 0 0 0 25 0 0 0 29 0 0 0 33 0 0 0 37 0 0 0 42 0 0 0 0 0 0 0 0 0 0 0]

This means that, for the 1st training sample (corresponding to the first row in ***train***), electrodes 5, 9, 13, 17 and 21 were occluded by 25 %, 29 %, 33 %, 37 % and 42 %, respectively.

***trainElev***

The file ***trainElev***contains a 77000x1 vector, in CSV format. Each row provides information on the belt elevation in degrees for the corresponding *train* measurements. For example, the 1st entry of ***trainElev***is 0.3030. This means that, for the 1st training sample (corresponding to the first row in ***train***), the belt was rotated by 0.303 degrees relative to the horizontal.

## Test data:

There are 110 test samples is one separate set of testing data, contained in files ***test***.

Note: DO NOT use any test data for training.

***test***

The file ***test*** contains a 110x6721 matrix, in CSV format. For each row, the first 6720 columns are the 6720 dimensions of the input vector, X. The last column is the output scalar, Y. Please see *Features* section below for an explantion of the 6720-dimension feature.

## Features:

There are 32 electrodes. For each heart volume, 224 unique combinations of current source/sink electrodes patterns are applied. For each pattern, 30 electrode voltages (32 minus the 2 involved in current sourcing/sinking) are measured. So, for each heart volume, there are 224\*30=6720 voltage measurements. This is why the X samples have 6720 dimensions.

The 6720 dimensions are grouped by pattern. X(1:30) are the voltage measurements from pattern #1. X(31:60) are the voltage measurements from pattern #2, etc.

The current source/sink pattern information is contained in the ***iipats.csv*** file. For instance, iipats(1, :) is [1, 2]. This means that for pattern #1, the source electrode is 1, while the sink electrode is 2. It also means that X(1:30) are measured from electrodes 3-32.

Similarly, iipats(80, :) is [8, 11]. This means that for pattern #80, the source electrode is 8, while the sink electrode is 11. Also, X(2371:2400) are measured from electrodes 1-7, 9-10, 12-32.

The electrode number of each of the 6720 voltages in X is contained in the ***elec.csv*** file. So, elec(1) is 3, and elec(2371) is 1.