**Assignment 2 Data Mining CSE 572**

Task 1: Extract meal data and no meal data from the data provided in Assignment 1 and another persons data provided in this assignment.

Meal data can be extracted as follows:

From the InsulinData.csv file search the column Y for a non NAN value. This time indicates start of meal consumption . Meal data comprises of a 2hr 30 min stretch of CGM data that starts from and extends to .

No meal data comprises of 2 hrs of raw data that does not have meal intake.

**Extraction strategy:**

Meal data

Start of a meal can be obtained from InsulinData.csv. Search column Y for a non NAN value. This time indicates start of a meal. There can be three conditions:

1. There is no meal from time to time . Then use this stretch as meal data
2. There is a meal at some time in between and . Ignore the meal data at time and consider the meal at time instead.
3. There is a meal at time , then consider the stretch from to as meal data.

No Meal data extraction:

Start of no meal is at time where is start of some meal. We need to obtain a 2 hr stretch of no meal time. So you need to find all 2 hr stretches in a day that have no meal and do not fall within 2 hrs of start of a meal.

Test Data:

The test data will be a matrix of size , where N is the total number of tests and 24 is the size of the CGM time series. N will have some distribution of meal and no meal data.

Note here that for meal data you are asked to obtain a 2 hr 30 min time series data, while for no meal you are taking 2 hr. However, a machine will not take data with different lengths. Hence, in the feature extraction step, you have to ensure that features extracted from both meal and no meal data have the same length.

Output format:

You have to output an vector of 1s and 0s, where if a row is determined to be meal data then the corresponding entry will be 1 and if determined to be no meal corresponding entry will be 0.

This vector should be saved in a Results.csv file.

Given: Meal Data and No Meal Data of subject 1 and 2

Ground truth labels of Meal and No Meal for subject 1 and 2

Todo:

1. Extract features from Meal and No Meal data
2. Make sure that the features are discriminatory
3. Trains a machine to recognize Meal or No Meal data
4. Use k fold cross validation on the training data to evaluate your recognition system
5. Write a function that takes one test sample as input and outputs 1 if it predicts the test sample as meal or 0 if it predicts test sample as No meal.

Grading:

50 points for developing a code in Python or Matlab that takes the given dataset, extracts meal and no meal data and trains a machine model

20 points for developing a code in Python or Matlab that implements a function to take a test input and run the trained machine to provide the class label as output

30 points will be evaluated on the accuracy, F1 score, Precision and Recall results obtained by your machine. This will be compared against class average to determine the final score.