Project 2: Machine Model Training

Purpose

In this project, you will use a training dataset to train and test a machine model. The purpose is to distinguish between meal and no meal time series data.

Objectives

Learners will be able to:

- Develop code to train a machine model.
- Assess the accuracy of a machine model.

Technology Requirements

- Python 3.6 to 3.8 (do not use 3.9).
- scikit-learn==0.21.2
- pandas==0.25.1
- Python pickle

Project Description

In this project, you will train a machine model to assess whether a person has eaten a meal or not eaten a meal. A training data set is provided.

Please watch the **three introductory videos on Project 2** before beginning.

- **Project 2: Machine Model Training Introductory Video 1 (https://canvas.asu.edu/courses/140907/files/62424119/download?wrap=1)**
- **Project 2: Machine Model Training Introductory Video 2 (https://canvas.asu.edu/courses/140907/files/62424121/download?wrap=1)**

- **Project 2: Machine Model Training Introductory Video 3 (https://canvas.asu.edu/courses/140907/files/62424124/download?wrap=1)**

Directions

Meal data can be extracted as follows:

- From the InsulinData.csv file, search the column Y for a non NAN non z indicates the start of meal consumption time tm. Meal data comprises a CGM data that starts from tm-30min and extends to tm+2hrs.
- No meal data comprises 2 hrs of raw data that does not have meal intak

Extraction: Meal data

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

1. There is no meal from time tm to time tm+2hrs. Then use this stretch as
2. There is a meal at some time tp in between tp>tm and tp< tm+2hrs. Ignc time tm and consider the meal at time tp instead.
3. There is a meal at time tm+2hrs, then consider the stretch from tm+1hr meal data.

Extraction: No Meal data

Start of no meal is at time tm+2hrs where tm is the start of some meal. We stretch of no meal time. So you need to find all 2 hr stretches in a day that h not fall within 2 hrs of the start of a meal.

Handling missing data:

You have to carefully handle missing data. This is an important data mining for many applications. Here there are several approaches:

1. Ignore the meal or no meal data stretch if the number of missing data pc greater than a certain threshold.
2. Use linear interpolation (not a good idea for meal data but maybe for nc
3. Use polynomial regression to fill up missing data (untested in this domai

Choose wisely.

Feature Extraction and Selection:

You have to carefully select features from the meal time series that are disc
meal and no meal classes.

Test Data:

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

Note here that for meal data you are asked to obtain a 2 hr 30 min time seri
meal you are taking 2 hr. However, a machine will not take data with differe
the feature extraction step, you have to ensure that features extracted from
meal data have the same length.

Output format:

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

- This vector should be saved in a "**Result.csv**" file.

Given:

- Meal Data and No Meal Data of subjects 1 and 2
- Ground truth labels of Meal and No Meal for subjects 1 and 2

Using Python, train a machine model to recognize whether a sample in the
represents a person who has eaten (Meal), or not eaten (No Meal). The tra
ground truth labels of Meal and No Meal for 5 subjects.

You will need to perform the following tasks:

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

Submission Directions for Project Deli

Deliverables:

- Two python files: 1) train.py and 2) test.py

- The train.py reads CGMData.csv, CGM_patient2.csv and InsulinData.cs extracts meal and no-meal data, extracts features, trains your machine t no-meal classes, and stores the machine in a pickle file (Python API picl
- The test.py reads test.csv which has the N x 24 matrix and outputs a Re N x 1 vector of 1s and 0s, where 1 denotes meal, 0 denotes no meal.
- Assume that CGMData.csv, CGM_patient2.csv and InsulinData.csv, Insı are all in your compilation and execution folder. Avoid using static paths.

Submission Guidelines:

- Please submit a zipped file containing train.py and test.py as "**yourfirstname_lastname_Project2.zip**".
- The submission space is located at the bottom of module 4 as "**Project 2: Machine Model Training Submission**".

Evaluation

Graders will evaluate your code as well as the accuracy of your results base and No Meal data that is not included in the training set.

- 50 points for developing a code in Python that takes the given dataset, e Meal data, and trains a machine model
- 20 points for developing a code in Python that implements a function to run the trained machine to provide the class label as output
- 30 points will be evaluated on the accuracy, F1 score, Precision, and Re by your machine.

Project-2-Files.zip (https://canvas.asu.edu/courses/140907/files/62424384?w (https://canvas.asu.edu/courses/140907/files/62424384/download?download_f

test.csv (https://canvas.asu.edu/courses/140907/files/62726471?wrap=1) ⬇ **(https://canvas.asu.edu/courses/140907/files/62726471/download?download_f**

Assignment 2 - Rubric			
Criteria	Ratings		
Extraction of meal data	10 pts Full Marks	0 pts No Execution	1(
Extraction of non-meal data	10 pts Full Marks	0 pts No Marks	1(
Classify and train data using K-fold cross validation.	30 pts Full Marks	0 pts No Marks	3(
Test the model and generate results file without any error	20 pts Full Marks	0 pts No Marks	2(
Calculate Accuracy of the training model	10 pts Full Marks	0 pts No Marks	1(
Calculate Precision, recall of the training model	10 pts Full Marks	0 pts No Marks	1(
Calculate F1 Score of the training model	10 pts Full Marks	0 pts No Marks	1(