DATASET

Our dataset can be easily got from the website of Kaggle competition, which consists of 3 parts: the training dataset (train.csv), the testing dataset (test.csv) and the professed classification of testing dataset (question.csv).

Both the training dataset and testing dataset provide 4 features of accelerometer recordings, including: UNIX time (T) and acceleration measured in g on x, y, z coordinate respectively (X, Y, Z). The training dataset provides a field of DeviceId, which is a unique Id of the device that generated the samples. For the testing dataset, each group of samples is labeled with a unique SequenceId. Each SequenceId is matched to a professed DeviceId in the questions.csv file, by which we can find the professed device that generated the sequence of accelerometer data in the test file.

Our task is to determine whether the accelerometer recordings in the test set (test.csv) belong to the proposed devices in the question set (questions.csv).