

Analitica Hackathon

Title: Market segmentation based on local customer activity

Team Members:

- Satish Palaniappan (tpsathish95@gmail.com)
- Siddharth G (gxrockinstyle@gmail.com)
- Vijayalakshimi MLS
(vijayalakshimisetthuraman@gmail.com)

BE CSE, SSN College of Engineering, 2015

Human Activity Recognition

- What is it?

Using smart phones (or smart devices) to monitor physical activity of human beings.

- Why is it necessary?

- Health care researchers would give a fortune to understand physical activity of groups of people
- To understand the sequence of common activities one performs throughout the day

The idea

- We use Human Activity Recognition and pattern mining to present a mirage of implementations where this would make a tremendous impact. The sectors we focus on:
 - Market Segmentation for targeted and focused product sales
 - Remote patient activity monitoring, for psychiatrists, doctors, healthcare professionals
 - Predicting the profession of an individual
 - Fitness app which gives feedback based on user activity patterns
 - A fun use case we present is, suggesting songs based on the mood of the person, which is reflected by the physical activity sequence they perform.

Our System

- To predict a sequence of humanly activities (from a basic set of 6) that any individual with a gyroscope, accelerometer & GPS (or to be more specific, having a smartphone) on them, is performing at any instant of time and there-by mining the relationships and associations between activity patterns of people.
- As our first step, we classify signals from the smartphone to indicate that the subject is in one of the following six activities:
 - SITTING
 - STANDING
 - WALKING
 - WALKING UP
 - WALKING DOWN
 - LAYING

Our System (cont.)

- With a classifier trained for physical activity recognition, our system collects similar but new data from GPS devices over time to form activity sequences (eg. STANDING – STANDING – SITTING – WALKING – WALKING – LAYING – LAYING – LAYING – LAYING – SITTING) and separate it on a user basis.

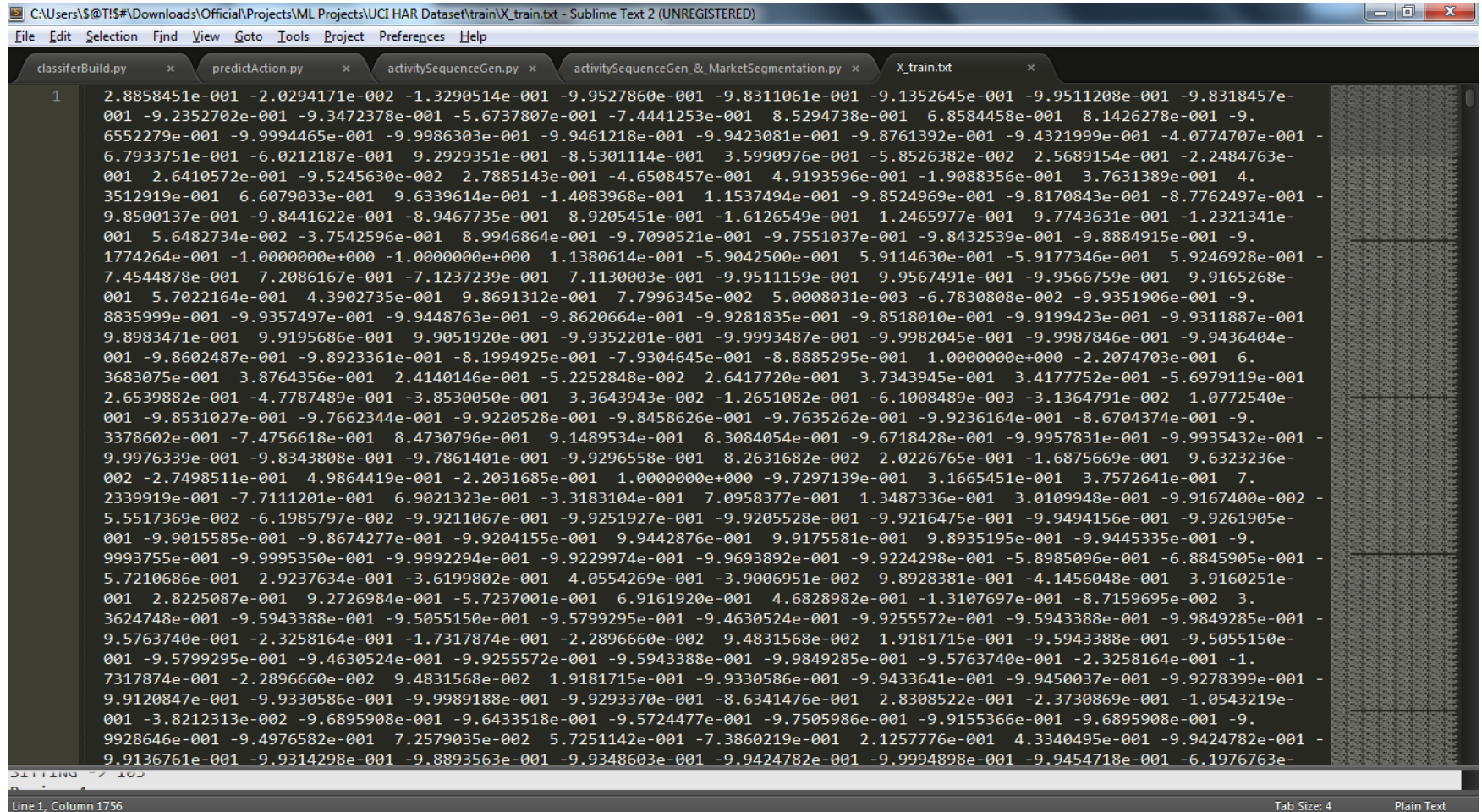
Data source

- University of California Irvine has a huge repository of datasets
- We used the dataset – Human Activity Recognition using Smartphones to train our classifier to identify physical human activity
- <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

DATASET

- The initial dataset was arrived at by analysing the gyroscope and accelerometer data from 30 people to label 6 activities.
- Features such as time domain acceleration, magnitude and a mirage of other features were extracted and Fast Fourier transformed to the frequency domain also.
- After some filters such as low pass filters were used to remove noise, the cleaned up dataset looks like this:
 - A matrix of filtered and normalized inertial signals (float numbers), of size $7352 * 561$ where each row represents a training example of 561 features, so in essence, we have 7352 training examples
 - A class label matrix of size $7352 * 1$, where each of the elements are from 1 – 6 representing one of the six classes

Dataset looks like this. This is a partial set of features of the first training example



```
C:\Users\T\Downloads\Official\Projects\ML Projects\UCI HAR Dataset\train\X_train.txt - Sublime Text 2 (UNREGISTERED)
File Edit Selection Find View Goto Tools Project Preferences Help

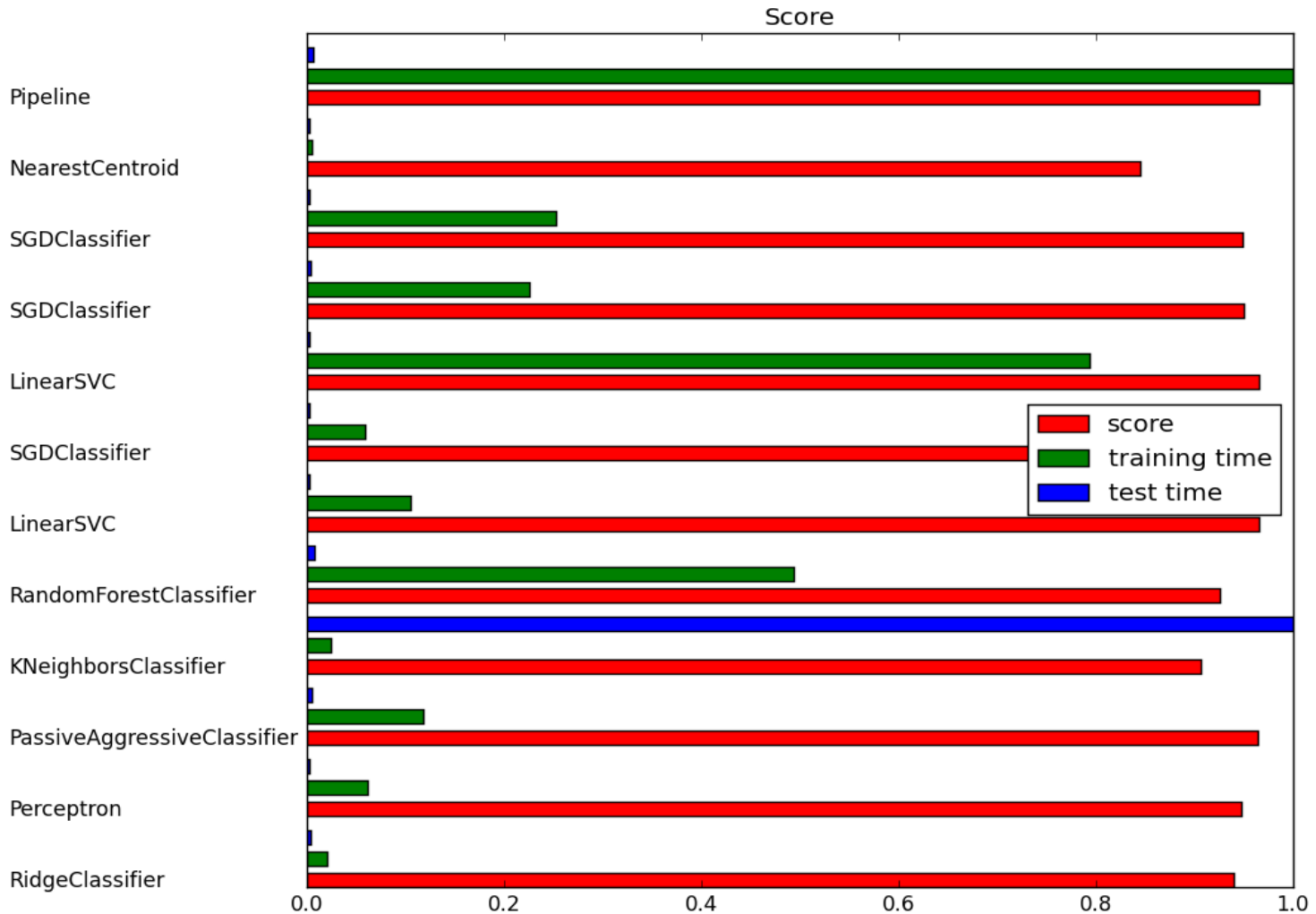
classifierBuild.py x predictAction.py x activitySequenceGen.py x activitySequenceGen_& MarketSegmentation.py x X_train.txt x

1 2.8858451e-001 -2.0294171e-002 -1.3290514e-001 -9.9527860e-001 -9.8311061e-001 -9.1352645e-001 -9.9511208e-001 -9.8318457e-
001 -9.2352702e-001 -9.3472378e-001 -5.6737807e-001 -7.4441253e-001 8.5294738e-001 6.8584458e-001 8.1426278e-001 -9.
6552279e-001 -9.9994465e-001 -9.9986303e-001 -9.9461218e-001 -9.9423081e-001 -9.8761392e-001 -9.4321999e-001 -4.0774707e-001 -
6.7933751e-001 -6.0212187e-001 9.2929351e-001 -8.5301114e-001 3.5990976e-001 -5.8526382e-002 2.5689154e-001 -2.2484763e-
001 2.6410572e-001 -9.5245630e-002 2.7885143e-001 -4.6508457e-001 4.9193596e-001 -1.9088356e-001 3.7631389e-001 4.
3512919e-001 6.6079033e-001 9.6339614e-001 -1.4083968e-001 1.1537494e-001 -9.8524969e-001 -9.8170843e-001 -8.7762497e-001 -
9.8500137e-001 -9.8441622e-001 -8.9467735e-001 8.9205451e-001 -1.6126549e-001 1.2465977e-001 9.7743631e-001 -1.2321341e-
001 5.6482734e-002 -3.7542596e-001 8.9946864e-001 -9.7090521e-001 -9.7551037e-001 -9.8432539e-001 -9.8884915e-001 -9.
1774264e-001 -1.0000000e+000 -1.0000000e+000 1.1380614e-001 -5.9042500e-001 5.9114630e-001 -5.9177346e-001 5.9246928e-001 -
7.4544878e-001 7.2086167e-001 -7.1237239e-001 7.1130003e-001 -9.9511159e-001 9.9567491e-001 -9.9566759e-001 9.9165268e-
001 5.7022164e-001 4.3902735e-001 9.8691312e-001 7.7996345e-002 5.0008031e-003 -6.7830808e-002 -9.9351906e-001 -9.
8835999e-001 -9.9357497e-001 -9.9448763e-001 -9.8620664e-001 -9.9281835e-001 -9.8518010e-001 -9.9199423e-001 -9.9311887e-001
9.8983471e-001 9.9195686e-001 9.9051920e-001 -9.9352201e-001 -9.9993487e-001 -9.9982045e-001 -9.9987846e-001 -9.9436404e-
001 -9.8602487e-001 -9.8923361e-001 -8.1994925e-001 -7.9304645e-001 -8.8885295e-001 1.0000000e+000 -2.2074703e-001 6.
3683075e-001 3.8764356e-001 2.4140146e-001 -5.2252848e-002 2.6417720e-001 3.7343945e-001 3.4177752e-001 -5.6979119e-001
2.6539882e-001 -4.7787489e-001 -3.8530050e-001 3.3643943e-002 -1.2651082e-001 -6.1008489e-003 -3.1364791e-002 1.0772540e-
001 -9.8531027e-001 -9.7662344e-001 -9.9220528e-001 -9.8458626e-001 -9.7635262e-001 -9.9236164e-001 -8.6704374e-001 -9.
3378602e-001 -7.4756618e-001 8.4730796e-001 9.1489534e-001 8.3084054e-001 -9.6718428e-001 -9.9957831e-001 -9.9935432e-001 -
9.9976339e-001 -9.8343808e-001 -9.7861401e-001 -9.9296558e-001 8.2631682e-002 2.0226765e-001 -1.6875669e-001 9.6323236e-
002 -2.7498511e-001 4.9864419e-001 -2.2031685e-001 1.0000000e+000 -9.7297139e-001 3.1665451e-001 3.7572641e-001 7.
2339919e-001 -7.7111201e-001 6.9021323e-001 -3.3183104e-001 7.0958377e-001 1.3487336e-001 3.0109948e-001 -9.9167400e-002 -
5.5517369e-002 -6.1985797e-002 -9.9211067e-001 -9.9251927e-001 -9.9205528e-001 -9.9216475e-001 -9.9494156e-001 -9.9261905e-
001 -9.9015585e-001 -9.8674277e-001 -9.9204155e-001 9.9442876e-001 9.9175581e-001 9.8935195e-001 -9.9445335e-001 -9.
9993755e-001 -9.9995350e-001 -9.9992294e-001 -9.9229974e-001 -9.9693892e-001 -9.9224298e-001 -5.8985096e-001 -5.8845905e-001 -
5.7210686e-001 2.9237634e-001 -3.6199802e-001 4.0554269e-001 -3.9006951e-002 9.8928381e-001 -4.1456048e-001 3.9160251e-
001 2.8225087e-001 9.2726984e-001 -5.7237001e-001 6.9161920e-001 4.6828982e-001 -1.3107697e-001 -8.7159695e-002 3.
3624748e-001 -9.5943388e-001 -9.5055150e-001 -9.5799295e-001 -9.4630524e-001 -9.9255572e-001 -9.5943388e-001 -9.9849285e-001 -
9.5763740e-001 -2.3258164e-001 -1.7317874e-001 -2.2896660e-002 9.4831568e-002 1.9181715e-001 -9.5943388e-001 -9.5055150e-
001 -9.5799295e-001 -9.4630524e-001 -9.9255572e-001 -9.5943388e-001 -9.9849285e-001 -9.5763740e-001 -2.3258164e-001 -1.
7317874e-001 -2.2896660e-002 9.4831568e-002 1.9181715e-001 -9.9330586e-001 -9.9433641e-001 -9.9450037e-001 -9.9278399e-001 -
9.9120847e-001 -9.9330586e-001 -9.9989188e-001 -9.9293370e-001 -8.6341476e-001 2.8308522e-001 -2.3730869e-001 -1.0543219e-
001 -3.8212313e-002 -9.6895908e-001 -9.6433518e-001 -9.5724477e-001 -9.7505986e-001 -9.9155366e-001 -9.6895908e-001 -9.
9928646e-001 -9.4976582e-001 7.2579035e-002 5.7251142e-001 -7.3860219e-001 2.1257776e-001 4.3340495e-001 -9.9424782e-001 -
9.9136761e-001 -9.9314298e-001 -9.8893563e-001 -9.9348603e-001 -9.9424782e-001 -9.9994898e-001 -9.9454718e-001 -6.1976763e-
001 -7.103
Line 1, Column 1756 Tab Size: 4 Plain Text
```


CLASSIFIERS

- We have a report of the accuracy of various classifiers we trained over this dataset in the file named 'report_classifiers.txt'.
- We have chosen a linear SVM (with L2 penalty) since it gave us a 97% accuracy at a fast training time.
- Comparison report is available on the next slide

Classifier comparison



The report for linear-SVC

=====

L2 penalty

Training:

LinearSVC(C=1.0, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, loss='l2', max_iter=1000, multi_class='ovr', penalty='l2', random_state=None, tol=0.001, verbose=0)

train time: 3.031s

test time: 0.047s

accuracy: 0.965

dimensionality: 561

density: 1.000000

classification report:

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

1	0.96	1.00	0.98	496
2	0.98	0.96	0.97	471
3	1.00	0.98	0.99	420
4	0.97	0.87	0.92	491
5	0.90	0.98	0.94	532
6	1.00	1.00	1.00	537

avg / total	0.97	0.97	0.96	2947
-------------	------	------	------	------

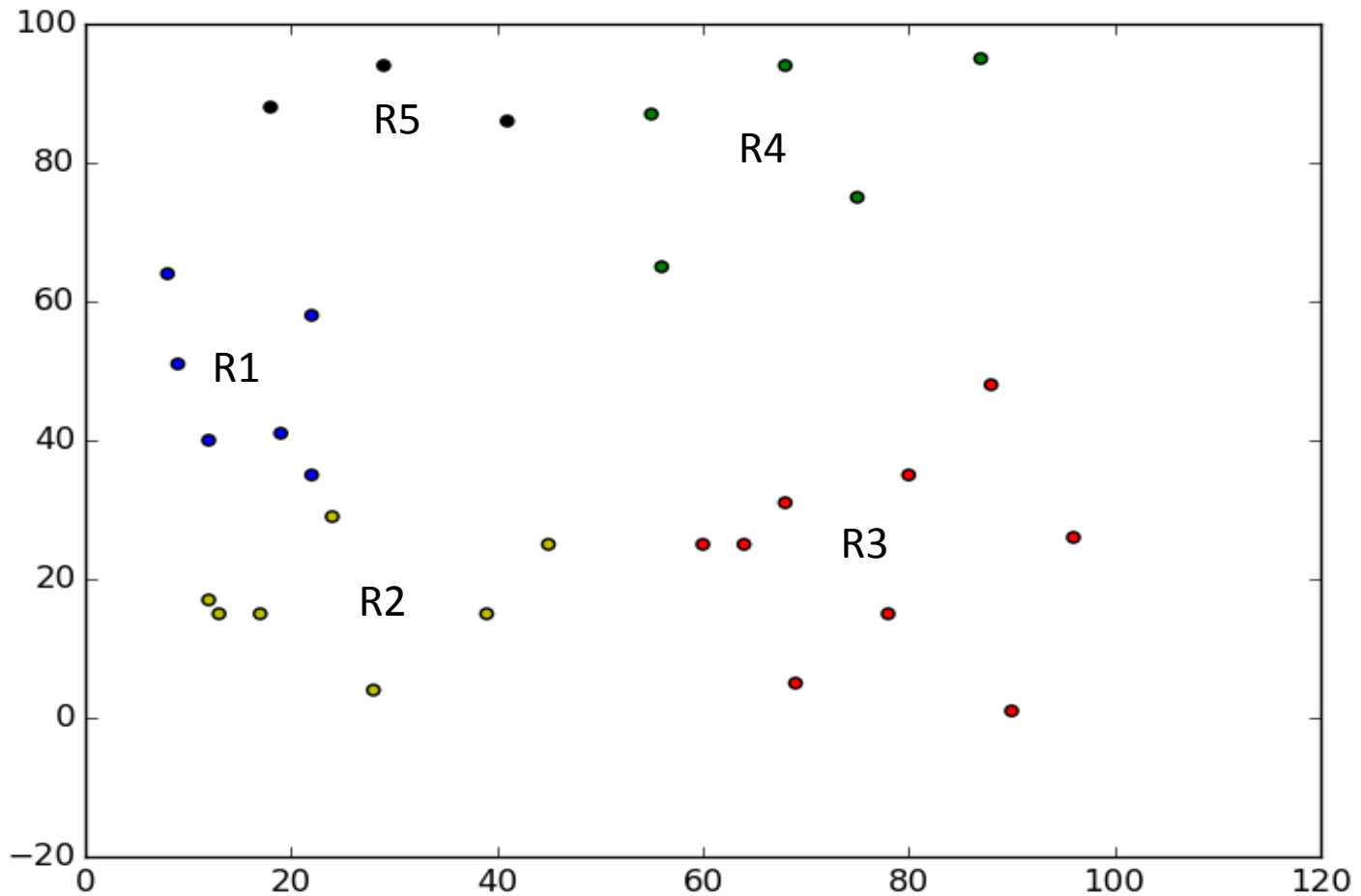
confusion matrix:

```
[[495  0  1  0  0  0]
 [ 20 450  1  0  0  0]
 [  2  5 413  0  0  0]
 [  0  3  0 429 57  2]
 [  1  0  0 11 520  0]
 [  0  0  0  0  0 537]]
```

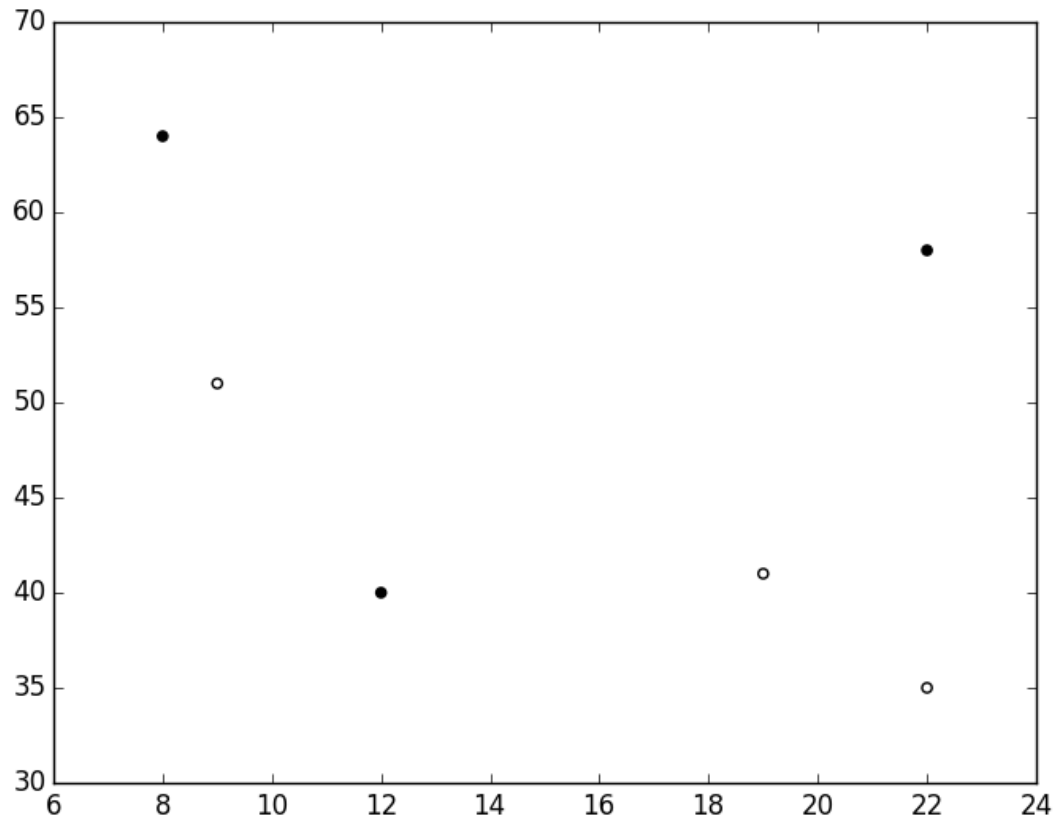
MARKET SEGMENTATION

- Then we input GPS locations along with training data of inertial signals for separate users where the signals are time ordered
 - Input of the form of the following tuple for ever user:
(GPS coordinates, (Training data in time sequence))
- Then we cluster the GPS coordinated based on geographical nearness.
- Each cluster is then analysed to produce activity report of populace

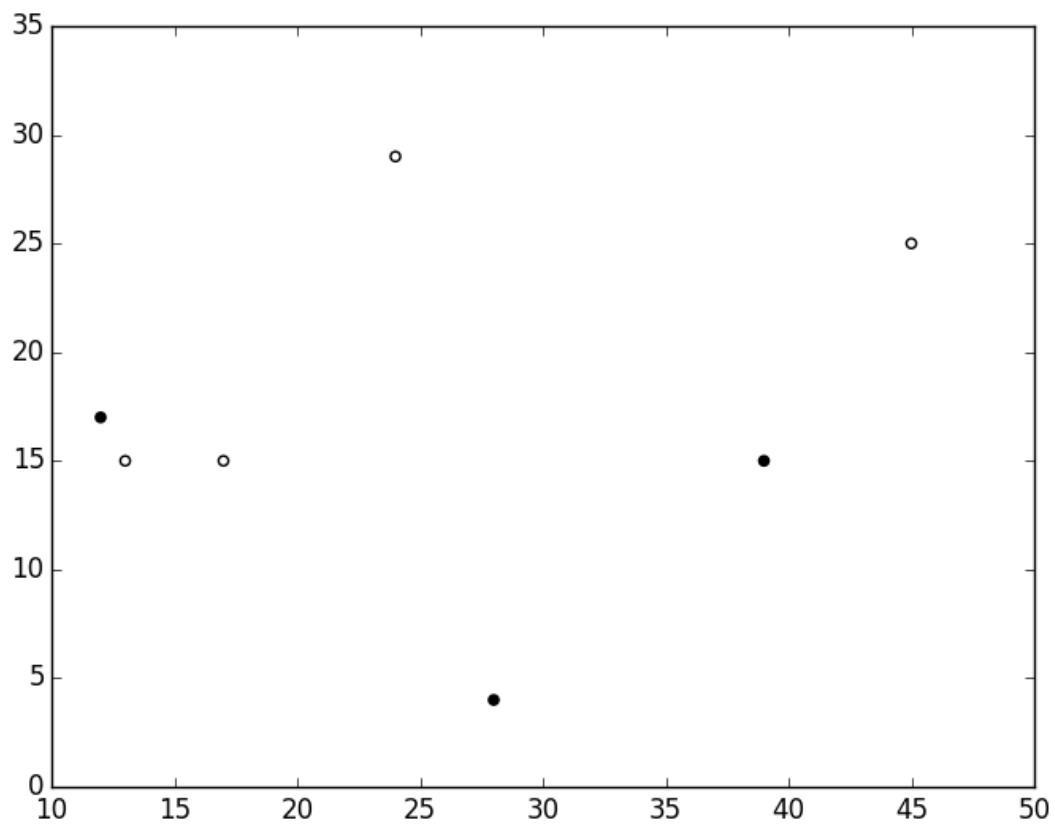
Geographic clusters



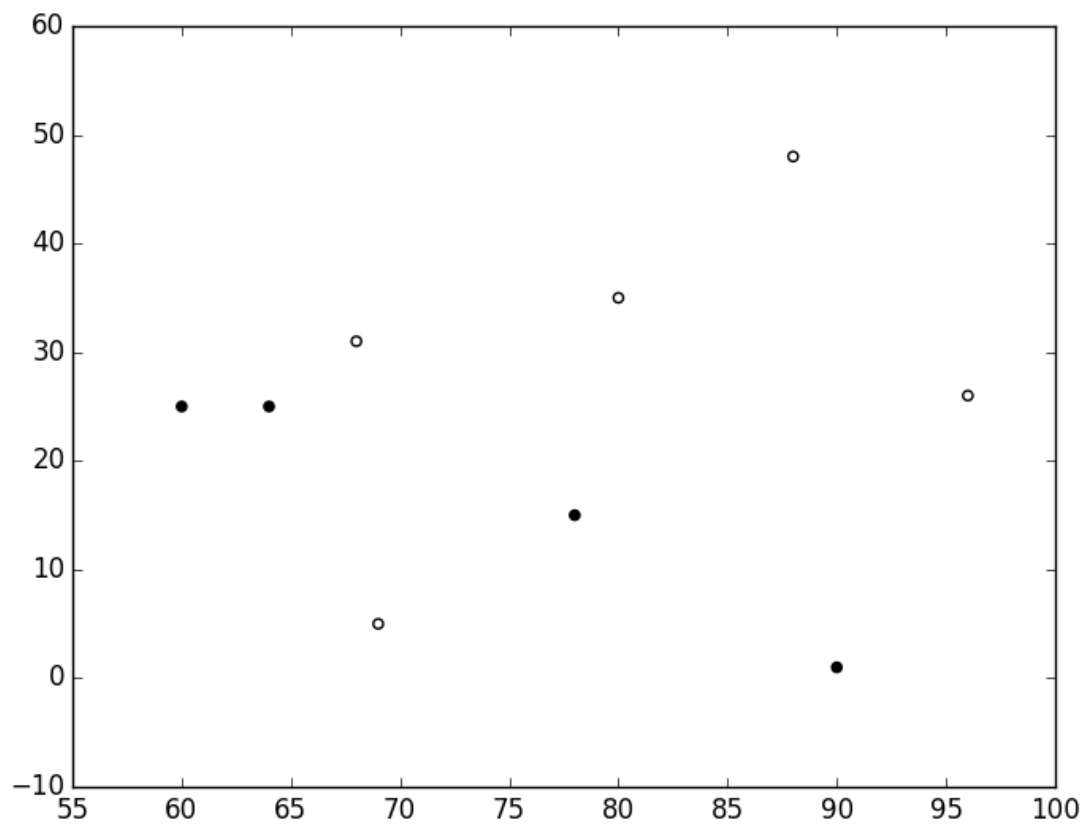
Activity analysis – the black and white labels here denote two different types of user activity patterns. The following image is of Region 1



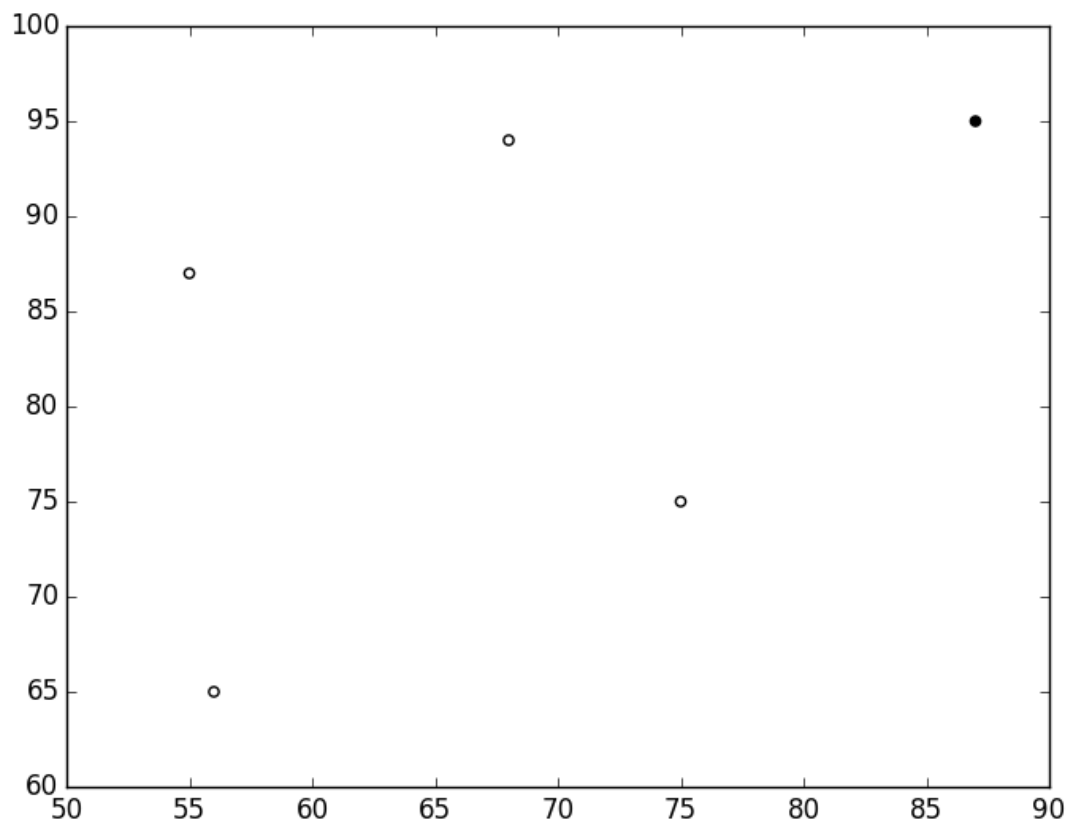
Region 2



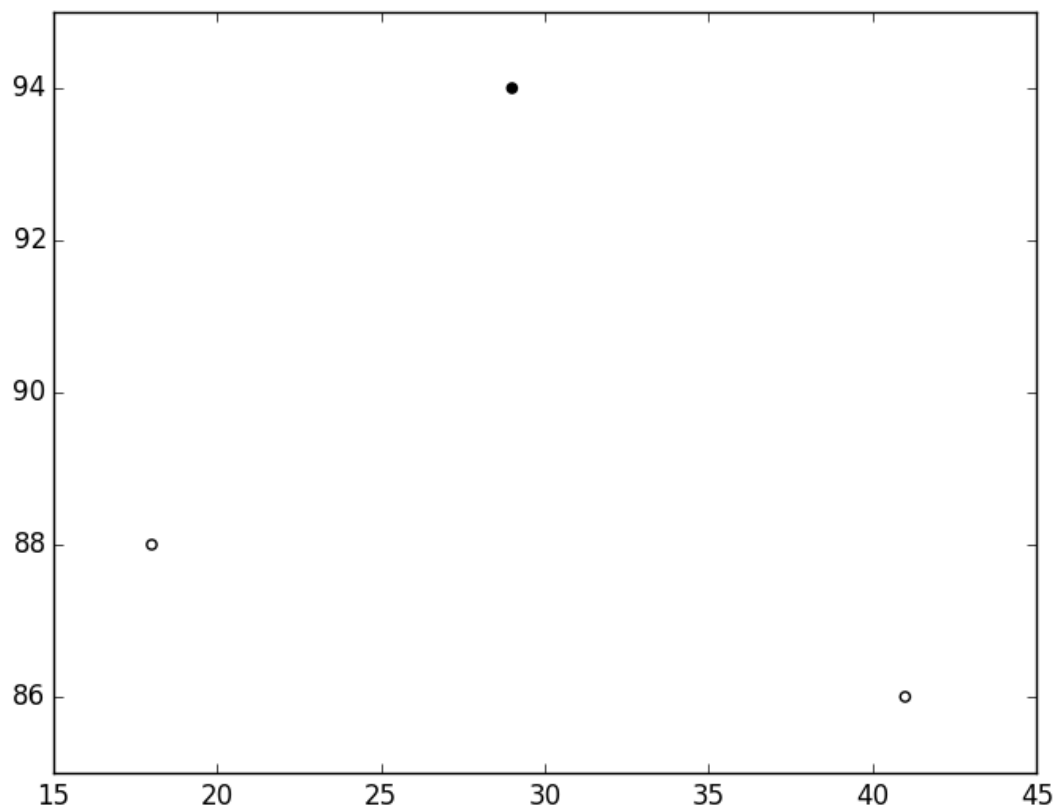
Region 3



Region 4



Region 5



OUTPUT

- We have made frequency analysis of the activities performed by individuals in a particular geographical cluster.
- This gives scope to target products to audience based on their activity patterns.
 - For eg, a location with high physical activity can be targeted to sell sports drinks
 - A location with low physical activity can be aimed at to sell fitness products, etc.
- The frequency analysis report for clusters in the images can be viewed at the file 'geo_frequency_output.txt'

What difference does this make?

- We can go a huge way to destroying obesity which is a plague infesting nations today.
- We can predict people's professions from their activity patterns and sell them products they need to enhance the overall health of the community
- A great but quirky use case would be monitoring mentally challenged patients in an asylum. There will be a single place to monitor activity of all patients and anomaly detection can be used to seek out erratic activity and render help to them really quickly.

Thank you for watching! 😊