Learning from Quantified Self Data

Roshan Vidyashankar
Stanford University
rvid@stanford.edu

Abstract

The Quantified Self is a movement to use technology to collect data on various aspects of a person's life like physical activity, nutrition, sleep, mental state, productivity etc. As mobile devices and sensors get smaller and cheaper, its getting easier to collect a lot of such information. The goal of this project is to see if we can find a correlation between this data and the person's mood.

1. Introduction

There has been an explosion of health tracking devices in the past few years. Devices like the Fitbit and Jawbone UP allow one to track many aspects of their life such as the total steps taken during the day, calories burned, sleep quality, heart rate etc. Health tracking apps have become a big part of all the major mobile operating systems (Android Fit[1] and Apple HealthKit[2]).
The Quantified Self movement however, is not new. People like Stephen Wolfram[3] and Nicholas Felton[4] have been tracking this kind of data for decades but mostly with a goal of using this data for interesting visualizations. In this project, I try to use some of this data to see how well it correlates with overall mood/well-being.

2. Dataset

A personal data set that was collected over a period of 1 year was used. It contains 115 data points, 1 data point per day. I collected part of this data by setting up a script that sends me an e-mail with a survey each night. It has information about my physical activity, productivity, mood, time spent on leisure activities(music) etc. Some information was also extracted from sources like Last.fm[5], Moves[6] and Github. The features extracted were:

- Workouts
- Productivity
- Music
- Sleep
- Walking-distance, steps, calories
- Running-distance, calories
- Cycling-distance, calories
- Total-calories-burnt

- Total-distance-travelled
- Music play count
- Hours spent at computer
- Number of git commits
- Mood

3. Methodology

3.1. First Look

To get an initial sense of what the data looked like, I tried Principal Component Analysis and Isometric Mapping. There were no obviously separable components.

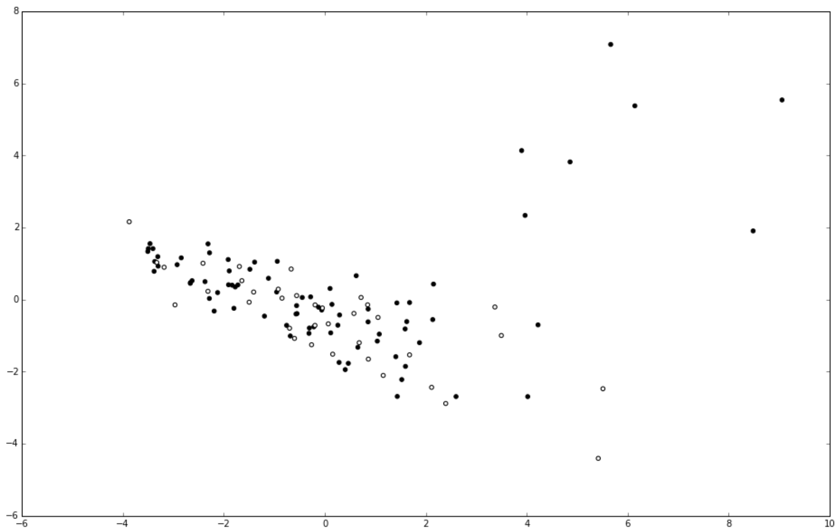


Figure 1 PCA

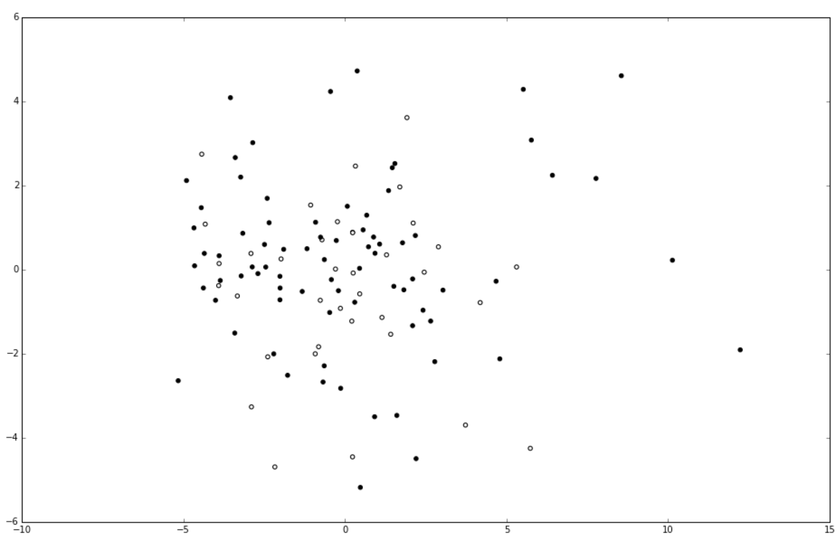


Figure 2 Isometric Mapping

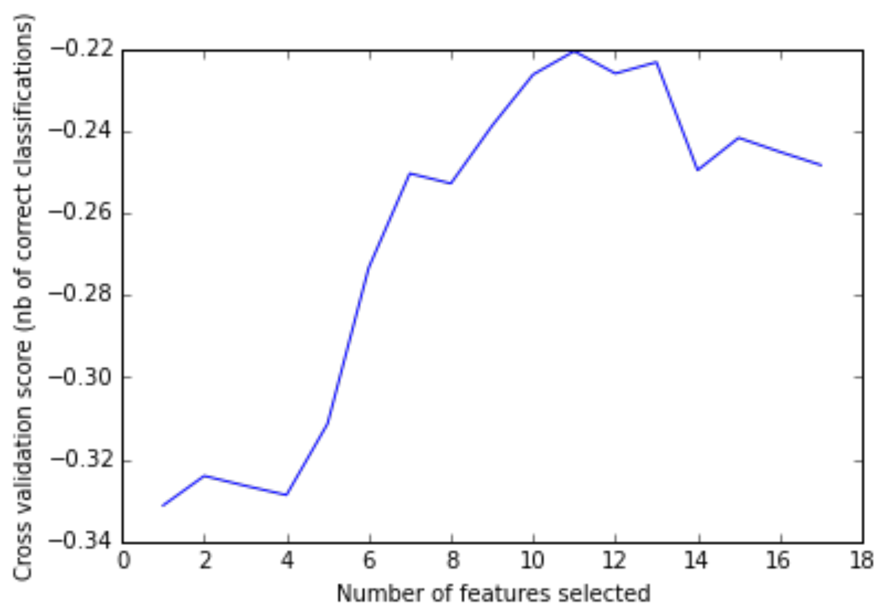
3.2. Regression

There were two main models that were used to regress over the mood based on the extracted features – Multiple Linear Regression and Support Vector Regression. Support Vector Regression was implemented with several different kernels – Linear, polynomial and RBF. The lowest Testing Error was achieved with Support Vector Regression with the RBF Kernel. Each of these models were tested with KFold cross validation.

Model	Training Error	Testing Error
Linear Regression	0.6692	1.429
SVR: Linear Kernel	9.393e-05	0.6292
SVR: Poly Kernel	0.00141	0.7418
SVR: RBF Kernel	1.011e-05	0.29811

3.3. Recursive Feature Elimination

Recursive Feature Elimination with Cross Validation was used to evaluate what features were most important in the feature set. The algorithm picked 11, the top 5 being - Sleep, Walking Distance, Total Calories Burnt, Git commits and Music Playcount.



4. Conclusion

The lowest error was achieved using SVMs with a RBF Kernel. However, this testing error had a very low corresponding training error which suggests over-fitting. It was interesting to see through Recursive Feature Elimination that features like Sleep, Walking Distance, Total Calories Burnt, Git commits and Music Playcount were strongly correlated with mood.

5. Future Work

There's certainly a lot of room for improvement. A large part of this is collecting more accurate data and minimizing self-logged data as much as possible. It will also be great to have data from multiple people - The Human Memome Project[7]'s dataset could be interesting to dig into. The biggest challenge is getting consistent data from multiple people over the same period of time. Another interesting approach would be to look at machine learning algorithms for time-series data to predict trends in mood.

6. References

[1] "Google Fit - Google Developers." 2014. 17 Nov. 2014 <https://developers.google.com/fit/>
[2] "Apple - iOS 8 - Health." 2014. 17 Nov. 2014 <https://www.apple.com/ios/whats-new/health/>
[3] "The Personal Analytics of My Life—Stephen Wolfram Blog." 2012. 17 Nov. 2014 <http://blog.stephenwolfram.com/2012/03/the-personal-analytics-of-my-life/>
[4] "Nicholas Felton | Feltron.com." 2002. 17 Nov. 2014 <http://feltron.com/>
[5] "roshanvid's Music Profile – Users at Last.fm." 2008. 17 Nov. 2014 <http://www.last.fm/user/roshanvid>
[6] "Moves - Activity Diary for iPhone and Android." 2013. 17 Nov. 2014 <https://www.moves-app.com/>
[7] "The Human Memome Project" <https://www.solveforx.com/moonshots/the-human-memome-project>