# Machine Learning Techniques applied to Heart Rate Variability

Final report for the Machine Learning course of the Technical University of Denmark

Pedro Filipe Emauz Madruga

2019-12-13

## Introduction

Heart Rate Variability (HRV) is a way to measure the variation in time in between each heartbeat (Campos (2017)). This variation is a measure of how the heart reacts to physical exercise, mental stress and heart diseases (Maritsch et al. (2019)), directly linked to an increased risk of mortality (Umetani et al. (1998)).

It has its origin on neurons from the parasympathetic, sympathetic nervous system and vagus nerve. Evidence suggests that HRV is impacted by stress (Kim et al. (2018)), specifically due to higher levels of stress resulting in a lower HRV (Altini (2017)).

While stress (and its causes and effects) is a known research topic, it's also more accessible due to the widespread usage of wearables that allow the collection of HRV data. The combination of the possibility of stress analysis from HRV and easy access to data, makes this the main focus of the present report, determining whether machine learning techniques can help minimalizing generalization errors.

# Part I

## Description of the data set

### Origin and introduction to the data sets

The data set is comprised of several different data sets, obtained from the Apple Watches of the author, over a course of around 2 years. In this time span, both the Apple Watch 3 and 4 were used and when referring to Apple Watch in this report, it pertains to both models - unless stated otherwise.

The data sets are:

- Activity Energy Burned: energy in kcal burned throughout different periods of the day.
- Apple Exercise Time: number of minutes within a given interval of time where exercise was tracked. Apple considers exercise as (Apple (2019a)):

  > "*[...] every full minute of movement that equals or exceeds the intensity of a brisk walk.*"

- Apple Stand Hour: number of times per hour that the subject has stood up.
- Flights Climbed: number of flights climbed within a given interval of time, measured several times during the day.
- Heart Rate: number of beats per minute measured every 5 seconds. The Apple Watch uses a technology called photoplethysmography, where (Apple (2019b))

  > "*[...]green LED lights paired with light-sensitive photodiodes to detect the amount of blood flowing.*"

- Mindful Session: time span (in intervals of seconds) where a meditation session was tracked.
- Heart Rate Variability (HRV): the Apple Watch computes the standard deviation of all normal sinus RR intervals over 24h (or SDNN). An RR interval is a beat-to-beat difference. When HRV is mentioned in this report, it refers to the SDNN values.
- Activity Summary: daily summary of some of the above mentioned features.

The data can be exported using an iPhone and the Export function inside the Health application. It originally it comes in a XML format and there is a number of tools available for conversion from XML to CSV format, one of which was used to do the conversion to CSV.

### Previous work with the data

Considering the lifetime of the Apple Watch, and assuming that the Heart Rate and HRV are to be included, the research is scarce and dispersed. There is research using data from Apple Watch, but falls into one of two categories:

improving existing measurements (Choksatchawathi et al. (2019)) or detection of various problems (as described in the introduction), some of which using Machine Learning and with success. In the latter category it was possible to verify that Machine Learning was used successfully to classify sleep-wake patterns (Chung et al. (2019)) and also detecting cardiovascular problems (Ballinger et al. (2018)).

However, the above mentioned cased are generic in a sense that only variations of the data set were used and for different subjects of study. In this project the data set pertains to one single user, with observations made over the course of 2 years.

### Primary machine learning modeling aim

This particular data set revolves around metrics of several different attributes. The principal challenge was to select a main attribute that was both relevant in terms of health impact, while leveraging the heart rate monitoring features that the Apple Watch provides, but also with enough observations so that it can be correlated with the remainder of different attributes monitored.

As mentioned in the introduction, HRV is an indicator of one's health status quo. Specifically, a lower HRV indicates both stress and cardiovascular potential problems, according to existing research. Considering other attributes monitored by the Apple Watch, such as the steps count and/or mindful session, it provides an interesting starting point to answer how is HRV affected by certain attributes.

Moreover, the data set with the HRV can be broken down into new interesting attributes. For instance, where is HRV higher or lower? Is it during work or after work hours? Is it during the week or the weekends? In which season of the year? Specifically, it can potentially be used for several different machine learning tasks, such as:

- Classification: how likely an HRV value belongs to a class of working hours or after-work hours. The original data doesn't have these features, thus its need to extract information from the data set. Specifically, this means that the existing HRV values (that are measured in various intervals of time), need to be split and grouped into two intervals of time (in a binary format): the first referring to an interval of time between 9am and 5pm (commonly referred as "working hours") and the period of non-working hours from 5pm to 9am. Data collected during sleeping time will be included because lower HRV (reflected as stress) can occur during the night. Nonetheless, most of measurements were made during wake time considering it was very rare that the Apple Watch was used during sleep time.
- Regression: so that it's possible to determine what's the estimated HRV value for the next day, based on the data from the previous days. The observations are made in periods more granular than a full day, but for this regression the average values of HRV will be used. There are two main reasons for this "grouping": the intervals where the observations happen

are not regular and don't have the same time span. For example, there could be several measurements in the morning in one day during a period of 5minutes and in another day the period could be shorter with just a few measurements made.

- Clustering: based on HRV measurements, determine if there was more or less physical activity (namely, whether the steps count and/or the flights climbed).
- Association mining: give certain features (such as step count and/or number of meditation sessions) are more likely to influence HRV values.

Despite being a rich data set (i.e. a reasonable number of observations made over a period of 2 years, with many attributes), it entails some limitations and issues. One of the main issues is the time span of observations, meaning that some observations were made several times during the day whereas others were not. Apple provides a summary of some of the attributes presenting observation values within a period of a day, but it's not for all attributes - namely "Active Energy Burned", "Apple Exercise Time" and "Apple Stand Hours" thus excluding HRV, flights climbed and step count.

Another issues is the definition of "Apple Exercise Time". As mentioned before, Apple describes it as anything more intense than a "brisk walk" without really defining what a brisk walk is. It's also not possible to determine which of the step count attribute counted as exercise time. The same for the flights climbed feature.

### Data attributes analysis

# References


Altini, Marco. 2017. "Heart Rate Variability: A (Deep) Primer." 2017. https://www.hrv4training.com/blog/heart-rate-variability-a-primer.

Apple. 2019a. "Get the Most Accurate Measurements Using Your Apple Watch." 2019. https://support.apple.com/en-gb/HT207941.

———. 2019b. "Your Heart Rate. What It Means, and Where on Apple Watch You'll Find It." 2019. https://support.apple.com/en-us/HT204666.

Ballinger, Brandon, Johnson Hsieh, Avesh Singh, Nimit Sohoni, Jack Wang, Geoffrey H. Tison, Gregory M. Marcus, et al. 2018. "DeepHeart: Semi-Supervised Sequence Learning for Cardiovascular Risk Prediction."

Campos, Marcelo. 2017. "Heart Rate Variability: A New Way to Track Well-Being." 2017. http://www-cs-faculty.stanford.edu/~uno/abcde.html.

Choksatchawathi, Tanut, Puntawat Ponglertnapakorn, Apiwat Ditthapron, Pitshaporn Leelaarporn, Thayakorn Wisutthisen, Maytus Piriyajitakonkij, and Theerawit Wilaiprasitporn. 2019. "Improving Heart Rate Estimation on Consumer Grade Wrist-Worndevice Using Physical Activity Level and Rolling Regression."

Chung, Yu-Min, Chuan-Shen Hu, Yu-Lun Lo, and Hau-Tieng Wu. 2019. "A Persistent Homology Approach to Heart Rate Variability Analysis with an Application to Sleep-Wake Classification."

Kim, Hye-Geum, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. 2018. "Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature." *Psychiatry Investigation* 15 (3): 235–45. https://www.ncbi.nlm.nih.gov/pubmed/29486547.

Maritsch, Martin, Caterina Bérubé, Mathias Kraus, Vera Lehmann, Thomas Züger, Stefan Feuerriegel, Tobias Kowatsch, and Felix Wortmann. 2019. "Improving Heart Rate Variability Measurements from Consumer Smartwatches with Machine Learning."

Umetani, Ken, Donald H Singer, Rollin McCraty, and Mike Atkinson. 1998. "Twenty-Four Hour Time Domain Heart Rate Variability and Heart Rate: Relations to Age and Gender over Nine Decades." *Journal of the American College of Cardiology* 31 (3): 593–601. https://doi.org/https://doi.org/10.1016/S0735-1097(97)00554-8.