Learning to predict sobriety to drive by using smartphone accelerometer data

# Motivation

Driving under the influence of alcohol causes many deaths each year. With this study, we want to create an easy and reliable way to detect if a person is still legally able to drive or not. We will predict a person’s sobriety, depending on their smartphone’s accelerometer data. We can then classify if the person is legally allowed to drive or not.

# Dataset

The dataset has 3 features:

* The X, Y and Z accelerations and their timestamps
* The trans-dermal alcohol concentration in the test person’s blood measured in g/dl and their timestamps
* A phone-type feature (Android or Apple)

This data was captured from 13 different people. The amount of alcohol in the blood is measured with SCRAM ankle bracelets in timeintervals of 30 minutes. This data has been cleaned by a filter and shifted 45 minutes backwards in time to match the real alcohol percentage in the blood more closely (because the received data is the alcohol leaving the person’s body through the skin 45 minutes after). More information on the processing of the dataset can be found in the paper about this dataset.

The phone-type will be ignored as we do not find this a useful feature. When we have extra time, we can further investigate whether this conclusion is valid or not.

# Method

In order to make a predictive model, we will slightly modify this dataset. We will use the data from the accelerometer as input and use the TAC data as an output. However, this output will not be a numerical representation of the percentage of alcohol in the person’s body. It will be changed into a Boolean to represent whether a person is legally able to drive or not. By doing so, we can make a classifier with two output classes.

We will make use of supervised learning to classify this dataset and try to make accurate predictions on new data. More precisely, we will use logistic regression. We can also create a neural network to determine the output, this would be our second choice.

# Reference

A paper has been published that used this dataset to make simular predictions. We will use this paper as a basis of our own research.

<http://ceur-ws.org/Vol-2429/paper6.pdf>