Final Project Report

Hao Zhang

2022-12-11

Abstract

The mobile health dataset consists of body movement and vital sign recordings of ten volunteers from various backgrounds while performing multiple physical activities. Sensors placed on the subjects' right wrist and left ankle were used to measure the movement of different body parts. Then,a binary logistic multilevel model is built to predict if a people is sitting or walking. After analysis, the predictors for this model are narrowed down to 3, which come from acceleration of right wrist. As a result, the coefficients of this model is convincing.

Introduction

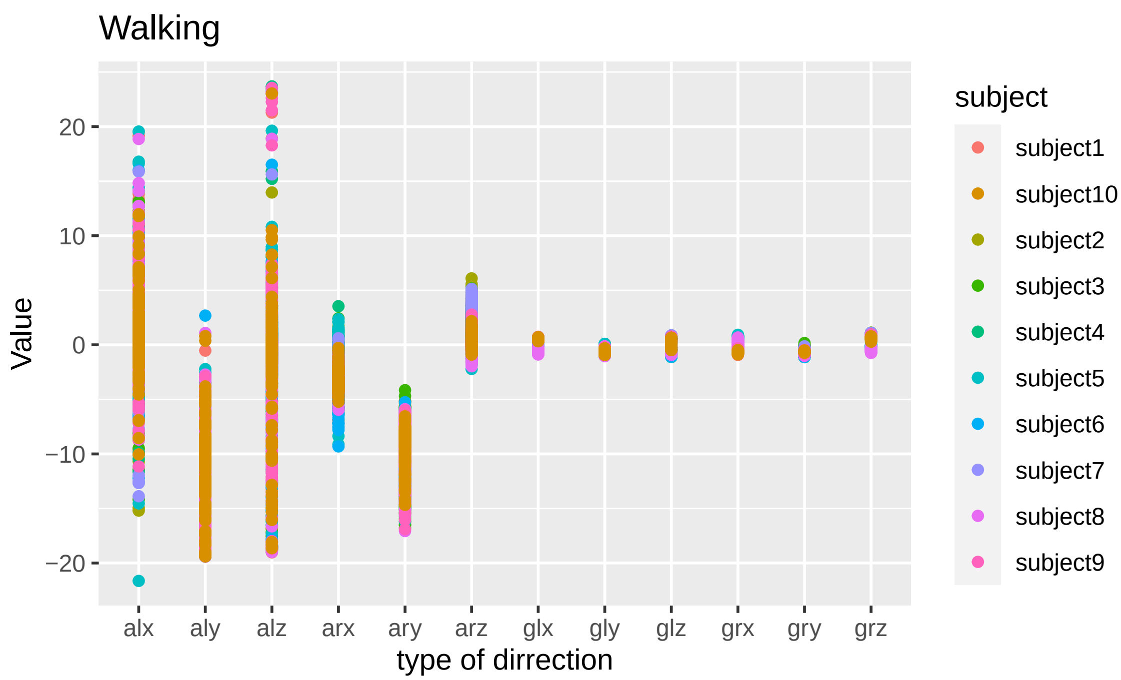
Modern sports watches all record some values when the body is exercising, such as the number of steps, heartbeat, etc., while some advanced watches will automatically detect whether the wearer is exercising. For example, when I walk faster with my Apple Watch, it will prompt me whether I am hiking and whether I need to record exercise information. When I sit for a long time, it will also remind me to stand up and move my body. So I'm curious how my watch sense actions. I found a very interesting data in this database, recording the body motion and vital signs of ten volunteers, each of them will wear sensors on the wrist and ankle to record three values during physical activity, respectively are acceleration, rotation angle and heartbeat. The data includes twelve activities, but I am only focusing on sitting and walking, hoping I can build a multilevel logistic model.

Method

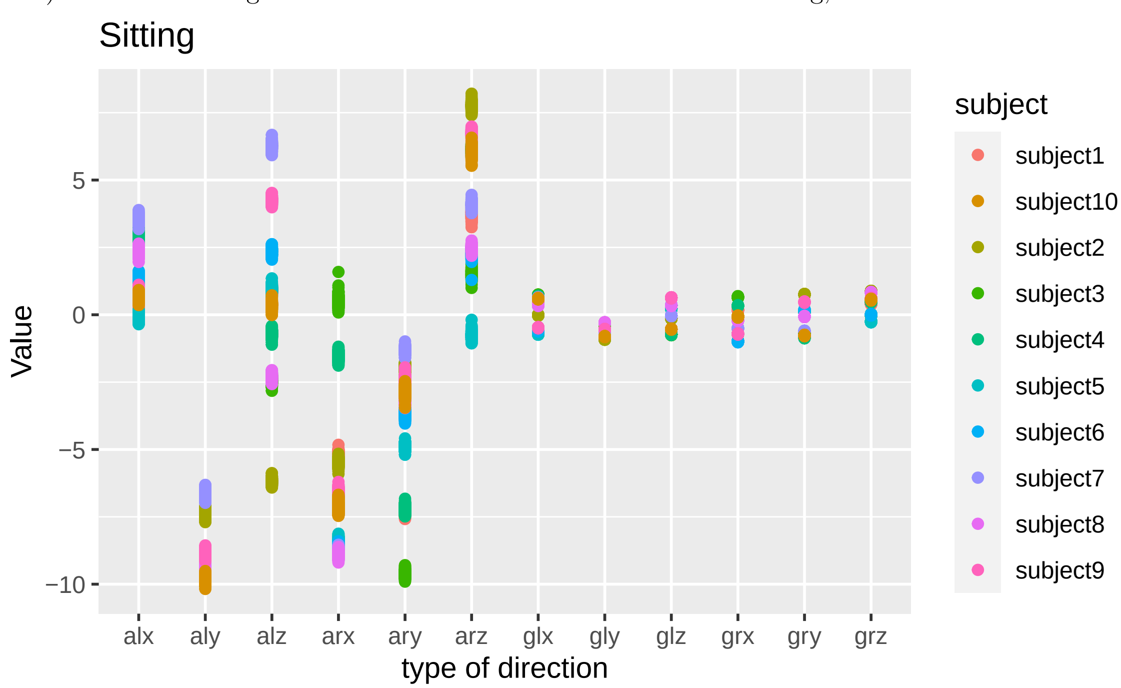
```
## [1] 0
```

Exploratory Data Analysis

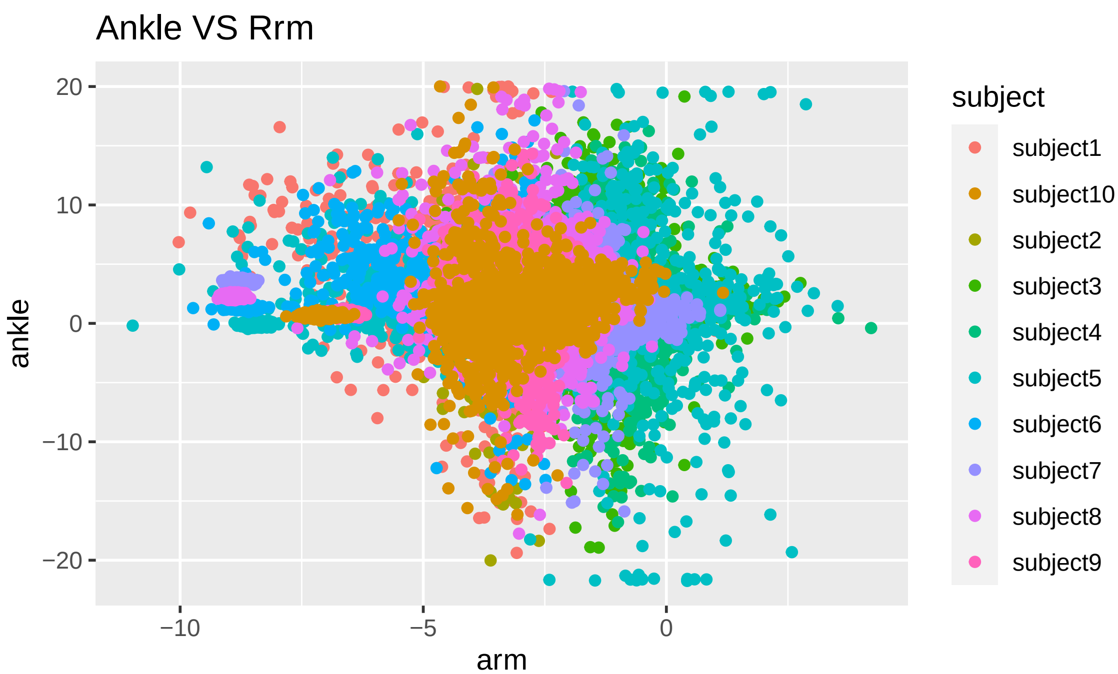
I have a data with 30720 observations of 14 variables. There are two type of sensors. One measure acceleration, while another one measure the rate of turn. Each sensor returns 3 numbers from 3 directions, x, y, and z. And each volunteers has two sensors on their the left ankle and right lower arm. Thus, there are total 12 predictors. Let's see how my data looks like.



This figure shows the data when 10 subjects are walking. And it seems that there are six predictors (rate of turn) that don't change a lot. It makes sense since when we're walking, we don't rotate our arms and legs.

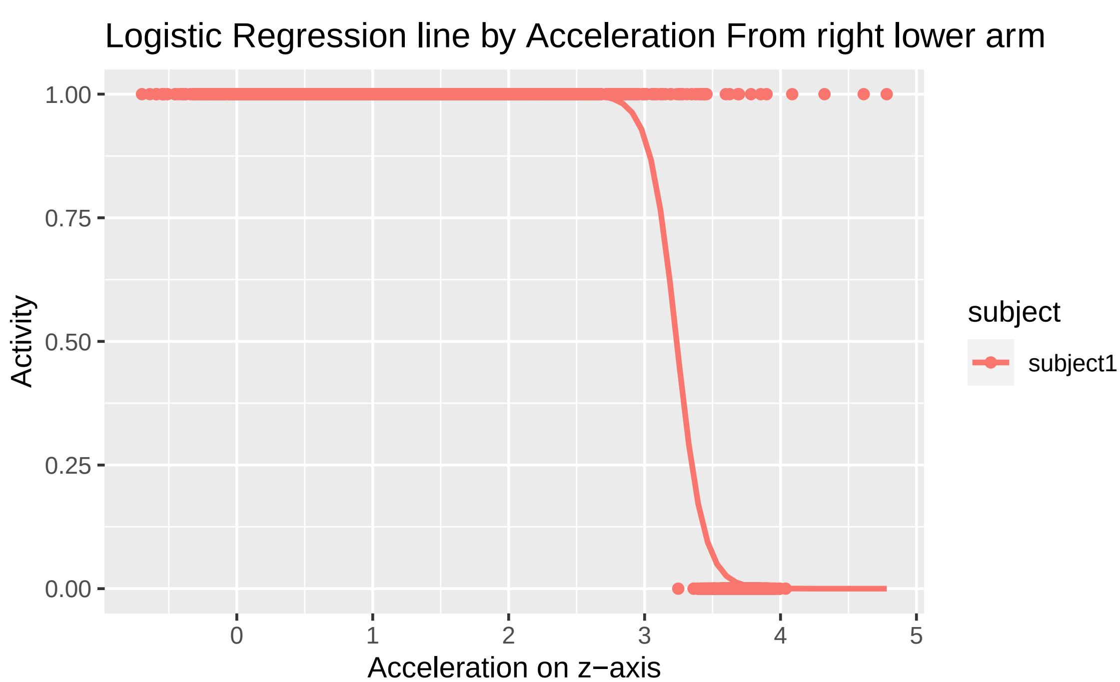


It shows the data when sitting. Obviously, different people have different sitting postures. It should be noticed that for both activities, comparing to acceleration, the rate of turn are pretty steady, hence I mainly focus on acceleration on ankle and arm and narrow down to 6 predictors. Next, I need to check if there is correlation

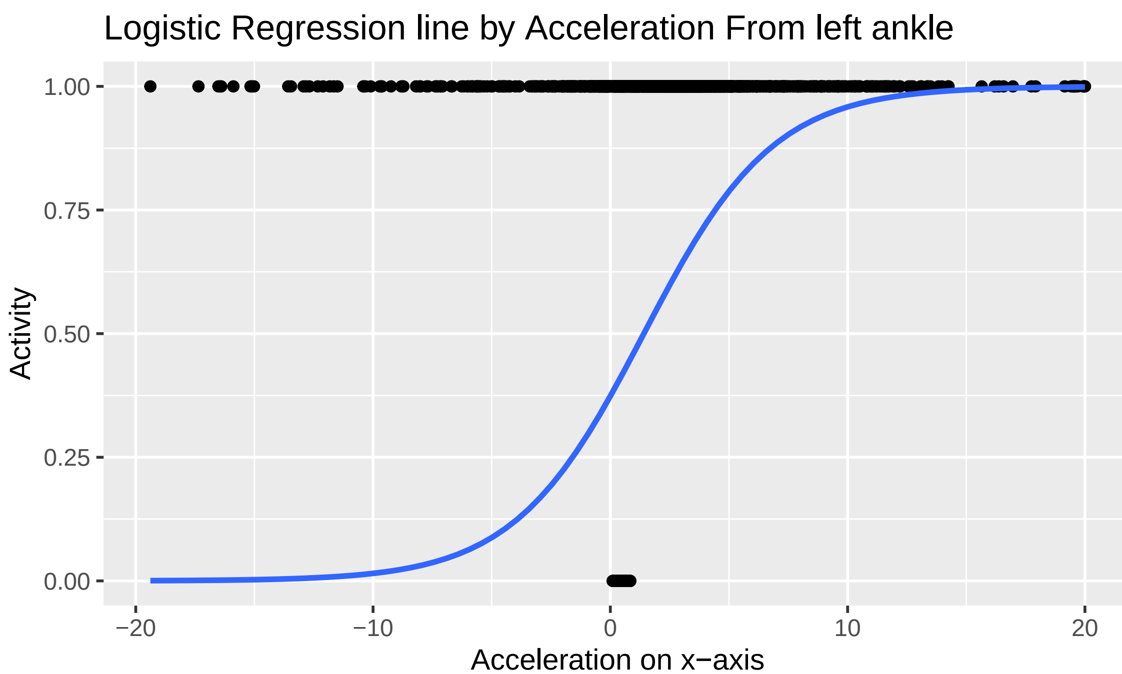


between ankle and arm.
It shows the comparison between left ankle and right arm. Each volunteer has his own acceleration, but overall, there is no correlation between acceleration and rate of turn. Also, I compared from other direction along y-axis and z-axis and I put plots on appendix.

Further, I need to make sure each predictor can fit a logistic regression in some way. Since they only presents as one direction, I cannot explain it by words but show each plots.



This figure shows a logistic regression line by acceleration from right lower arm. Here I only take subject1 as



an example.

Figure 5 show a logistic regression line by acceleration from left ankle. From this plot, alx (acceleration of left ankle along x-axis) is not a good predictor.

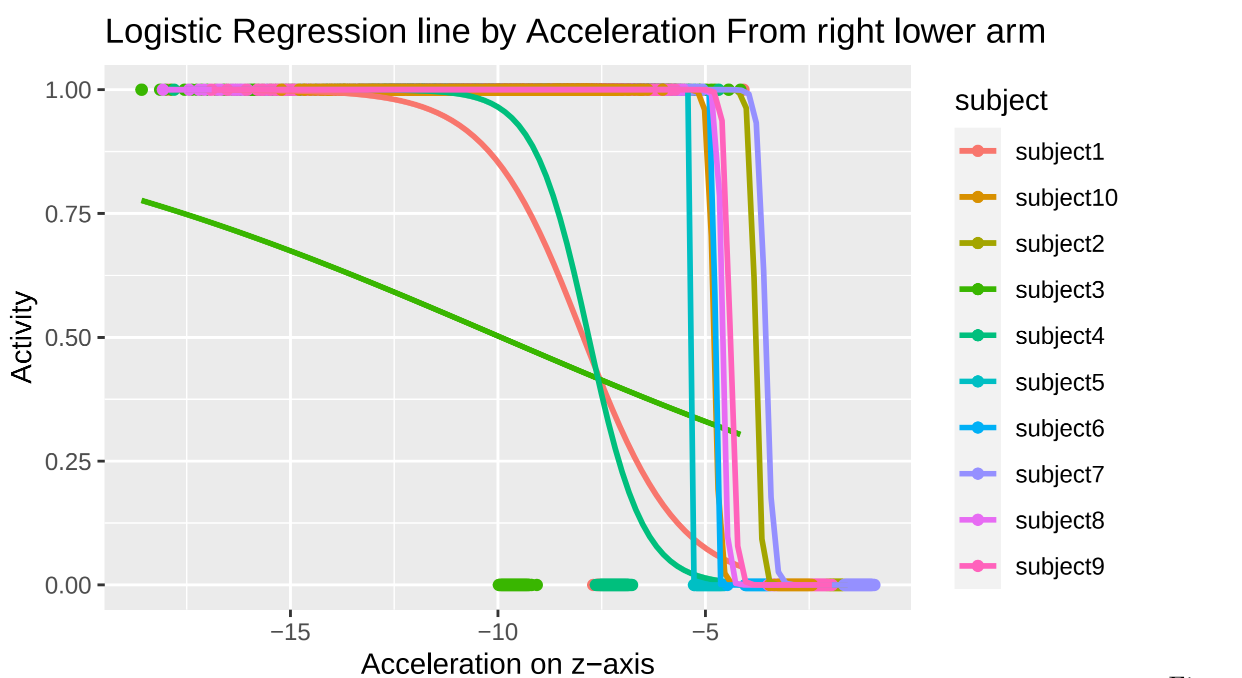


Figure 6 also shows the logistic regression line, just like previous two plots, but it concludes 10 volunteers. Overall, the maximum slop don't vary from subject to subject a lot, which means we don't have to consider ary (acceleration of left ankle along y-axis) as a random effect. Although it doesn't converge, we can consider it as a reference.

Model

Here is the function I built.

```
model1<-glmer(Activity~1+arx+ary+arz+(1+arx+arz|subject),data=data,family=binomial(link="logit"))
```

Here is the summary of fixed effects model and all variable are considered as statistically significant at $\alpha\ = 0.5$ level.

```
          Estimate Std. Error z value Pr(>|z|)
```

(Intercept) 20.5627767 0.0005316 38679 <2e-16 ***arx 2.6606881 0.0005316 5005 <2e-16*** ary -0.8578468 0.0005316 -1614 <2e-16 ***arz -3.0882063 0.0005316 -5809 <2e-16***

And here is the summary of random effects model. both arx and arz vary a lot.

```
## $subject
##             (Intercept)         arx         arz
## subject1    -0.49768591 -2.89661877 -5.68956042
## subject10    0.02292822 -0.00481417 -0.02625356
## subject2    -3.41501271 -0.76560012  0.83482030
## subject3   -26.98751311 -8.25865326  1.98132385
## subject4    -1.19425392 -1.46784178 -2.20092175
## subject5    -8.10099810 -0.21866458  5.28352862
## subject6    -3.01393391  0.30447057  2.77499296
## subject7    -0.19286079  0.22974663  0.61505136
## subject8     1.71626280  0.64218373  0.11729645
## subject9    -0.73730081 -0.13000360  0.25338408
##
## with conditional variances for "subject"
```

Result

Interpretation:

Subject2: The fixed effect formula is: $logit(p) = 20.56 + 2.66 * arx - 0.86 * ary - 3.09 * arz$

add random effect, we get:$logit(p) = 20/54 + 2.66 * arx - 0.86 * ary - 3.12 * arz$

Notice, all three parametters are negative.

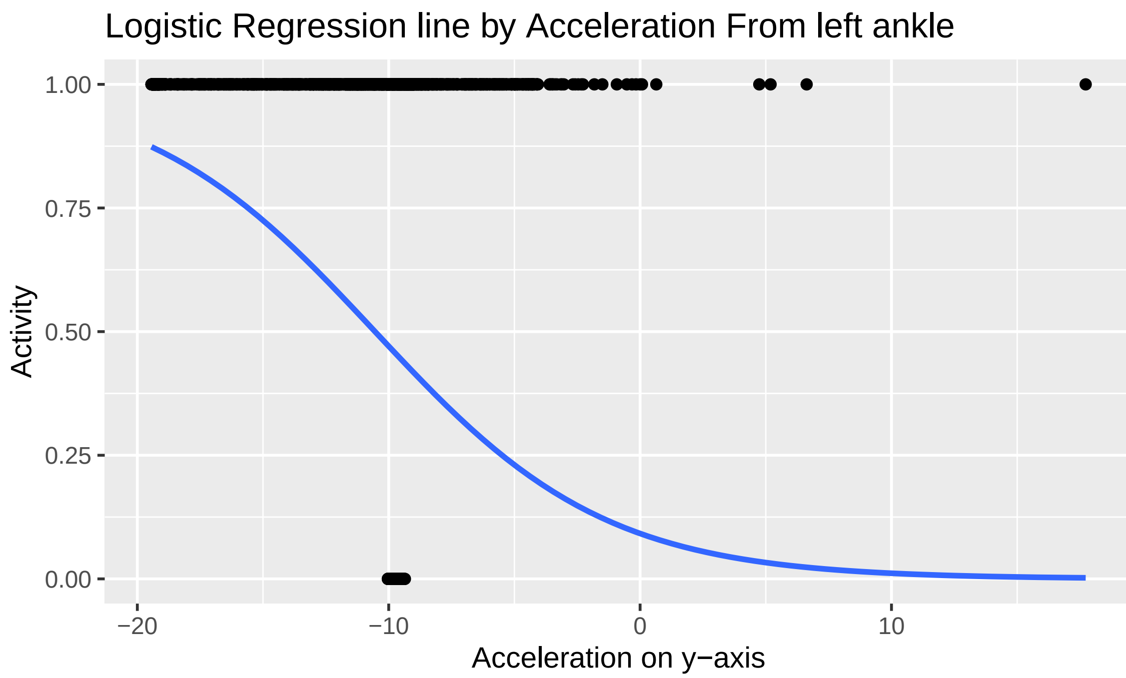
for every increase 1 unit of arx will result in a 2.66 increase in logit(p) If log(p/1-p) increases by 2.66, that means that p/1-p will increase by exp(2.66)=14.3. This is a 1330% increase in the odds of being walking. It makes sense that the coefficient of arx is much higher than other two, because the acceleration on x-axis is much easier to tell if a people is sitting or walking; waving arms toward (walking) changes a lot along x-axis.

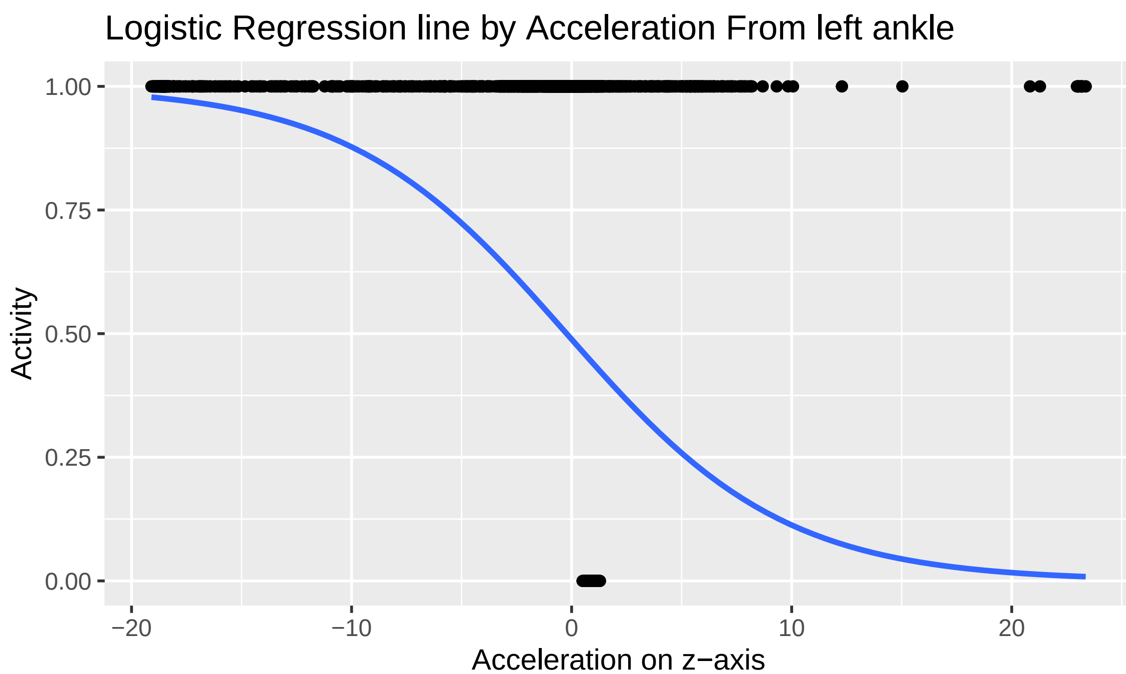
for every increase 1 unit of arz will result in a -3.12 increase in logit(p) If log(p/1-p) increases by -3.12, that means that p/1-p will increase by exp(-3.12)=0.044. This is a 95.6% decrease in the odds of being walking. It is noticed that the acceleration of waving arms along z-axis don't change a lot, while, when sitting there, arms are more likly to move up and down.

Discussion

The result of this report is that I only need to know the sensor data on my arm to tell whether I am sitting or walking. The result is exactly what I expected and behaved similarly to what my watch exhibited. Further, if other activities are also taken into consideration, we need a multinomial logistic model to analyze more kinds of activities; also we have to reconsider predictors.

Appendix




Logistic Regression line by Acceleration From left ankle
1.00
0.75
0.50
0.25
0.00
Activity
−20
−10
0
10
20
Acceleration on z−axis