2021/01/20 Ex.2

Marco Scarpelli

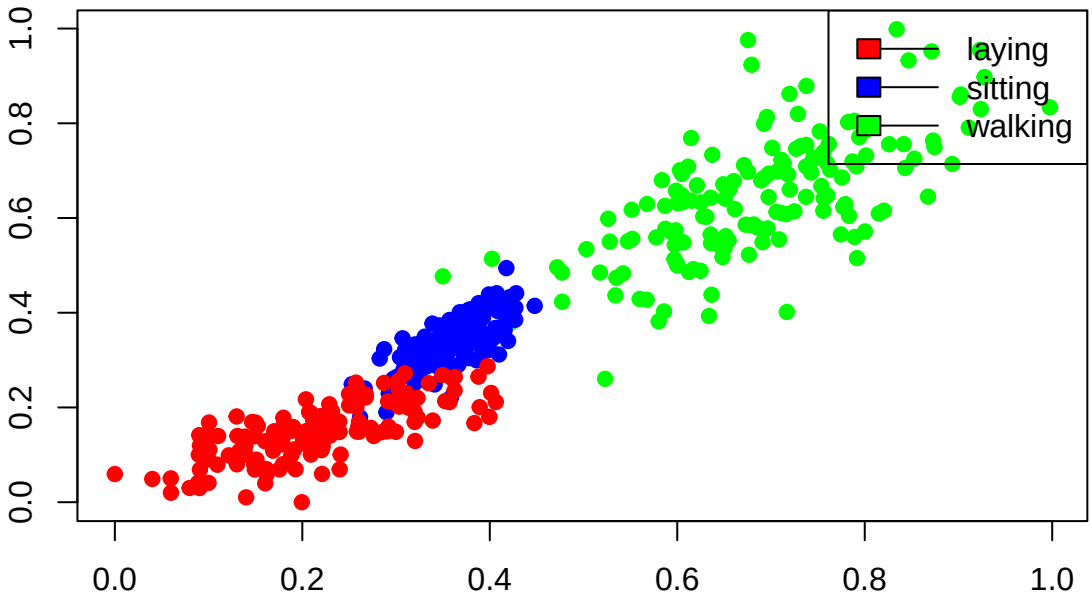
05 giugno, 2024

Dataset exploration

```
##   accel gyro activity
## 1  0.57 0.63  walking
## 2  0.71 0.40  walking
## 3  0.74 0.73  walking
## 4  0.61 0.55  walking
## 5  0.59 0.51  walking
## 6  0.60 0.50  walking
```

```
## [1] 450   3
```

We will import and jitter each variable separately, with a standard deviation that is 1% of the variable's mean.



Point a

Assumptions

We want multivariate normality within the groups:

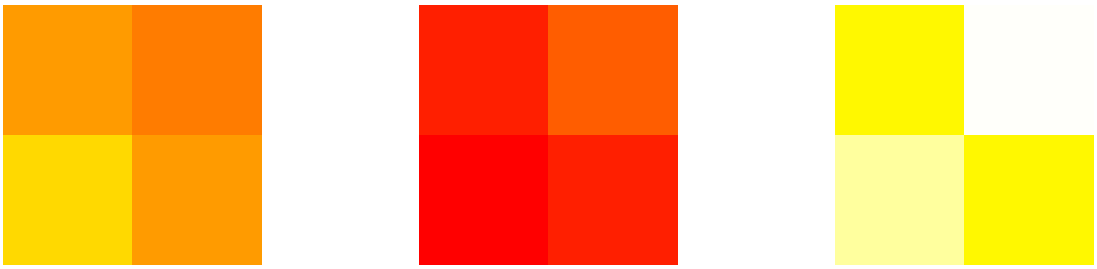
```
##            Test        HZ  p value MVN
## 1 Henze-Zirkler 0.5250285 0.624299 YES
```

```
##            Test        HZ   p value MVN
## 1 Henze-Zirkler 0.6162811 0.4349427 YES
```

```
##            Test        HZ   p value MVN
## 1 Henze-Zirkler 0.3902264 0.8862219 YES
```

All tests report normality. Let us check how the covariance structure is, to decide whether we should use LDA or

Cov. SA **Cov. SB** **Cov. SB**



QDA:

We will print them to check numerically:

```
##             accel        gyro
## accel 0.007147523 0.003800963
## gyro  0.003800963 0.003714738
```

```
##             accel        gyro
## accel 0.001489654 0.001623229
## gyro  0.001623229 0.002947237
```

```
##            accel       gyro
## accel 0.01307362 0.01021242
## gyro  0.01021242 0.01766216
```

To use LDA, we should observe that no value on the diagonal of each matrix is greater than 4 times larger than the corresponding diagonal values on the other matrices. In this case, we cannot use LDA.

QDA

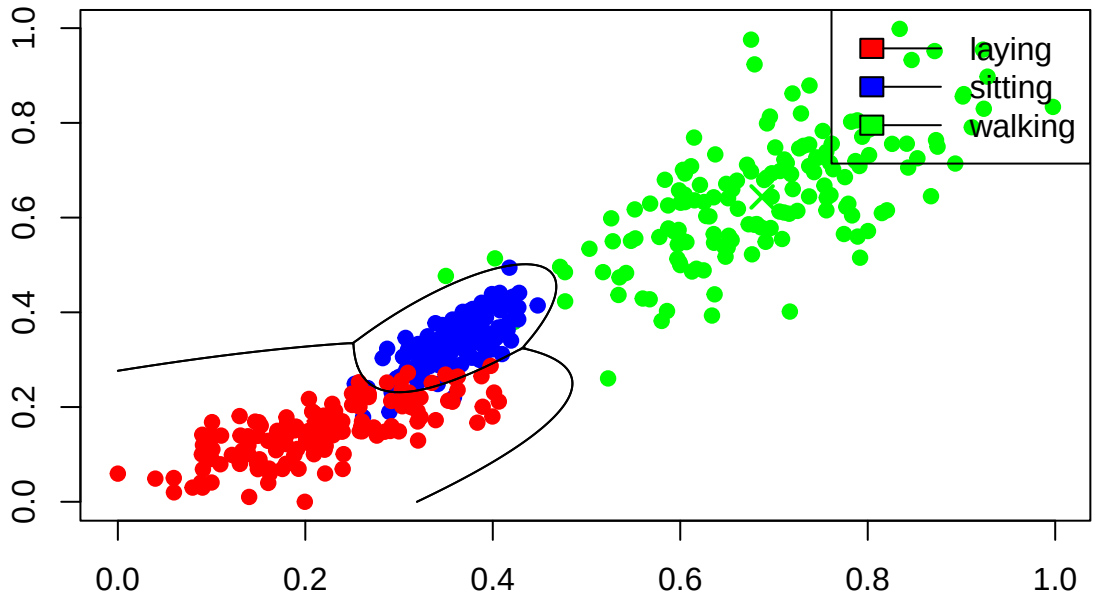
Means within groups:

```
##     accel      gyro
## 0.2147706 0.1472968
```

```
##     accel      gyro
## 0.3586588 0.3375291
```

```
##     accel      gyro
## 0.6873116 0.6435986
```

Classification regions:



Point B

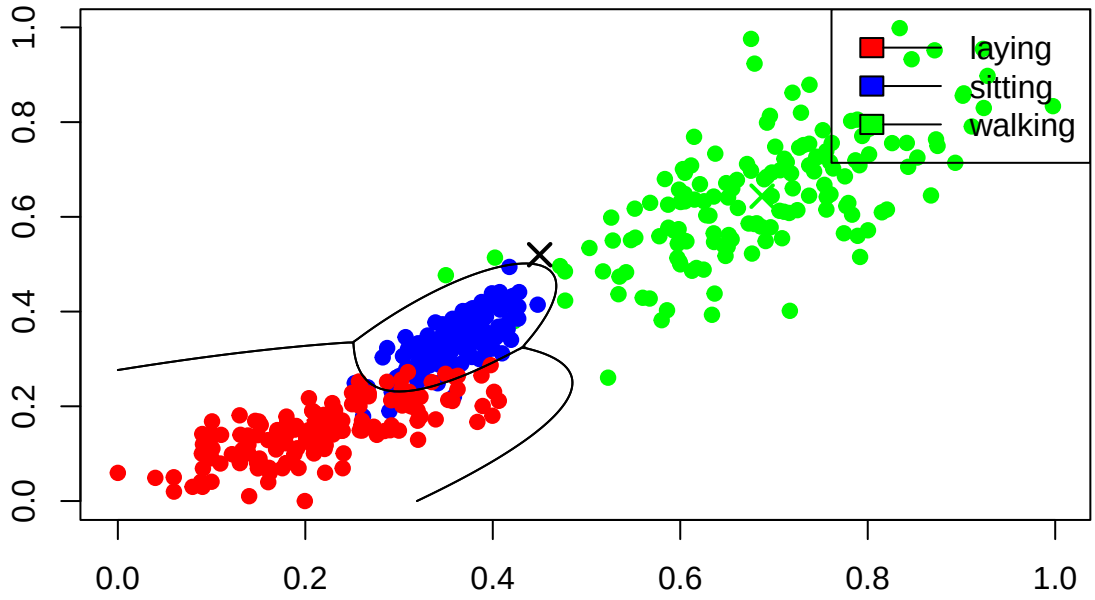
```
## [1] 0.03777778
```

Point C

```
## $class
## [1] walking
## Levels: laying sitting walking
```

```
##
## $posterior
##            laying   sitting   walking
## [1,] 7.522171e-09 0.2263664 0.7736335
```

The new point is identified by a black cross:



Point D

We classify the dataset and report the resulting table:

```
##          group
## data.knn  laying sitting walking
##   laying     144       6       1
##   sitting      6     144       3
##   walking      0       0     146
```

The APER is:

```
## [1] 0.03555556
```

The two classifiers perform similarly, with the KNN one being slightly worse.