Data Analysis on Human activity prediction based on smartphone data sets

sun

Problem

How to use sensor signal data from smart devices to predict human activity:

walking,sitting,lying,walking stairs,running and so on

Data

Data: UCI

http://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions

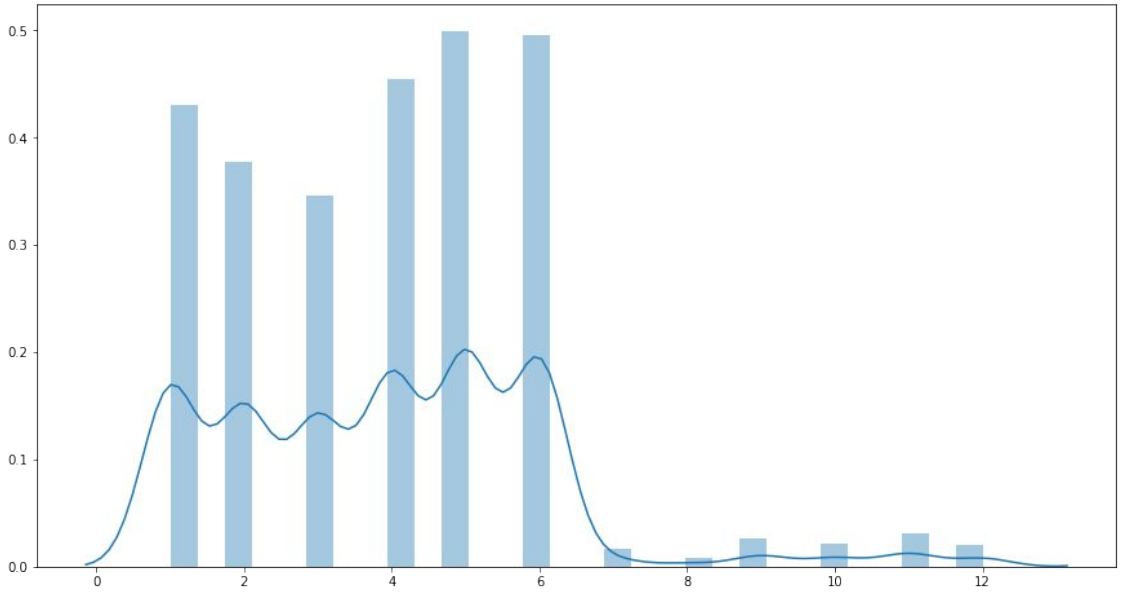
Sensor signals from smartphones of a group of 30 people performed a protocol of activities(such as walking,sitting ,lying and walking stairs) were collected and pre-processed.

Data Analysis

- 1 . Load Data
- 2.Missing Data
- 3.Unbalance Data
- 4.Outlier
- 5.Correlation
- 6.Null hythophis

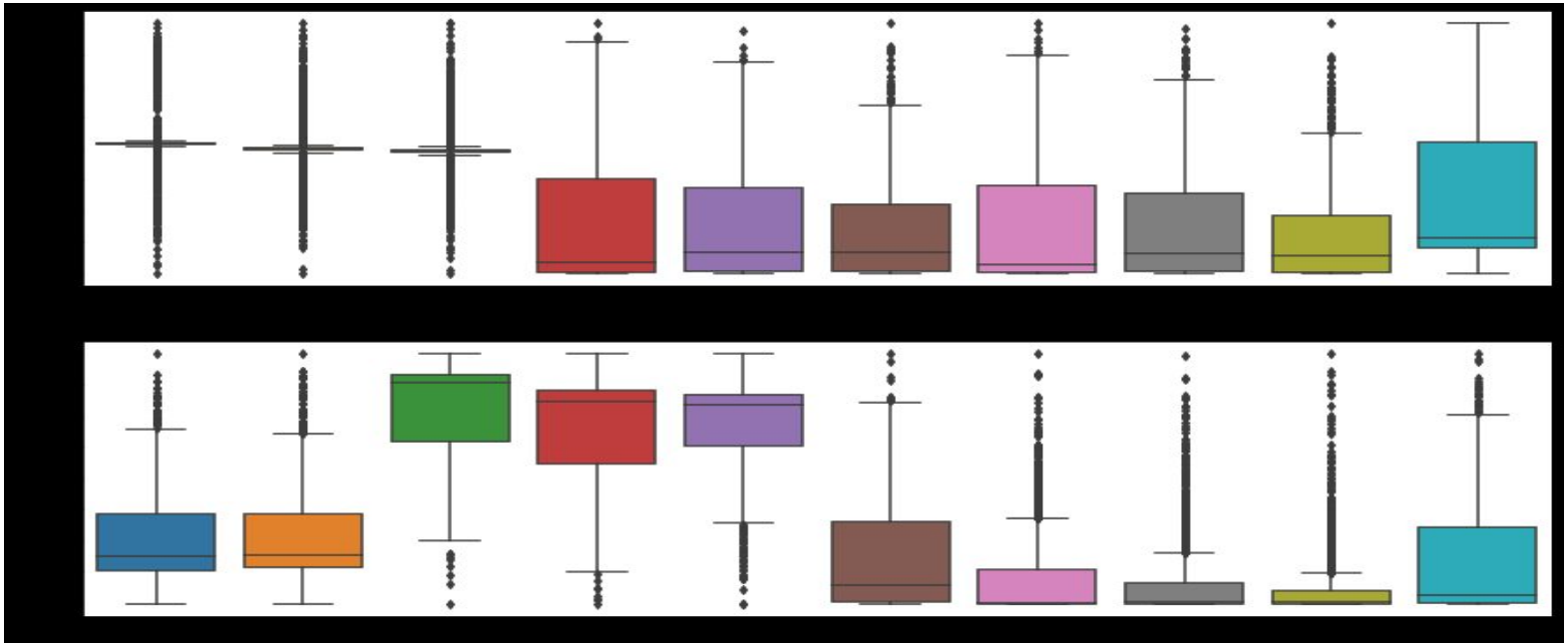
Balance Data

The barplot shows our dataset has 6 major classes and 6 minor classes, it is unbalanced, so we applied SMOTE to resample the data



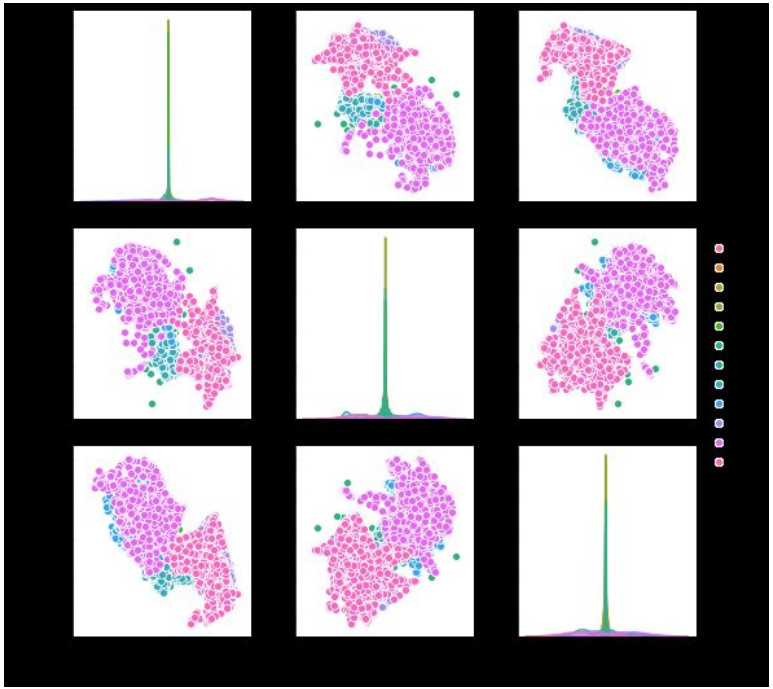
Outliers

According to the boxplot , we can see that there are a lot of features having outliersFor example feature 8 has a lot of larger outliers,we refill the outlier with 75% value ,feature 13 has smaller outliers ,we will refill those with 25% value



Correlation

The results of plot pairs of the first 3 numeric variables shows some variables are correlated with each other , for example feature 0 and feature 1



Null hythophis

Hythophis:There is no significent difference between below two sets:Set a) feature 2 value larger than 0.5, y belong to class 4 Set b) feature 2 value less than 0.5, y not belong to class 4

Result:The pvalue is 7.363006721969925e-25 < 0.05, so the null hythophis is rejected, there is a big chance if feature 2>0.5 ,it will belong to class 4.

Summary

- No missing data
- All numeric fields
- Unbalance dataset
- A lot of Outliers
- Pairbox shows some features are correlated
- When feature 2>0.5, there is a big chance ,the final class will be 4.