STA 135 Project Outline

1/24/2014

1 MEMBERS OF OUR GROUP

Yilun Zhang	1st year Master of Biostatistics
Dan Zhang	1st year Master of Statistics
Yu Liu	1st year Master of Statistics

2 BRIEF DESCRIPTION

The data comes from an experiment aiming to human activity recognition using Samsung smartphones. The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz.

The first 561 columns are measurements from Samsung Smartphone and the subject variable records which subject this measurements from and the activity variable indicates the activity performing.

The raw data and description can be accessed from here:
http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones.

The data we use is processed to be easily loaded in R(can be downloaded from https://spark-public.s3.amazonaws.com/dataanalysis/samsungData.rda). I also commit it

with some other things on github:https://github.com/lunge111/STA135PROJECT (The data is under Rworkspace). I will keep works updated on github once we figure out something on this project.

3 PROBLEMS WE FOCUS ON

3.1 Correctness of recording activity.

We can use classification to determine wether all observations are classified correctly. If the proportion of wrongly discriminated observations is big, it's reasonable to say that it's hard to determine human activities by samrtphone measurements.

3.2 Clustering all observations.

We will clustering observations into 6 clusters and see wether the measurements do provide information on human activity.

3.3 Explore a mechanism to determine human activity by respective smartphone measurements.

We probably use some statistical methods like regression and ANOVA to determine relationship(if any) between (part of) smartphone measurement and human activities. This would let smartphone to know what human's doing, thus the phone can adjust some setting depending on human activity. Turn off the rotation function when people's lying, for example.

4 STATISTICAL TECHNIQUES MAY BE USED

Classification: to find wether there exists wrongly discriminated observations.

Clustering: to cluster observations.

Dimension reduction: to expedite calculation of R because there are as many as 561 variables

of data.

Some other applicable methods we learn in this course.