Project

CSCI 6515 - 2017

Dr. Amilcar Soares
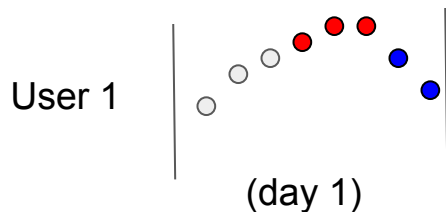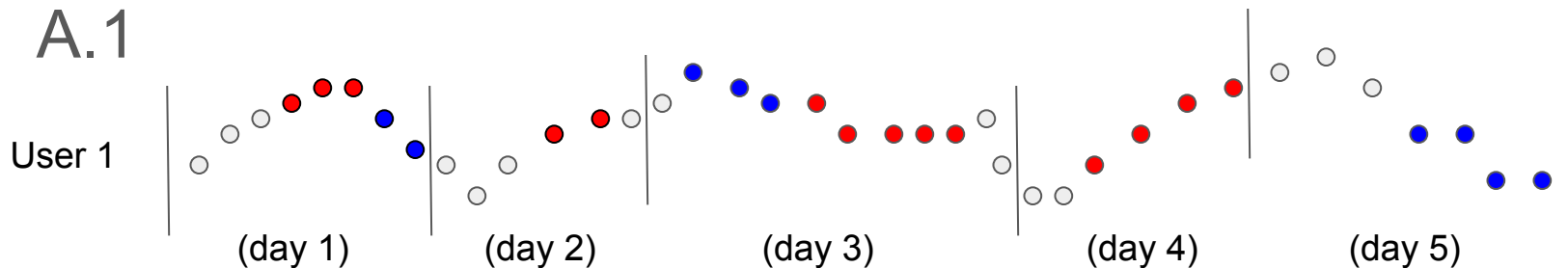
Institute
for Big Data
Analytics

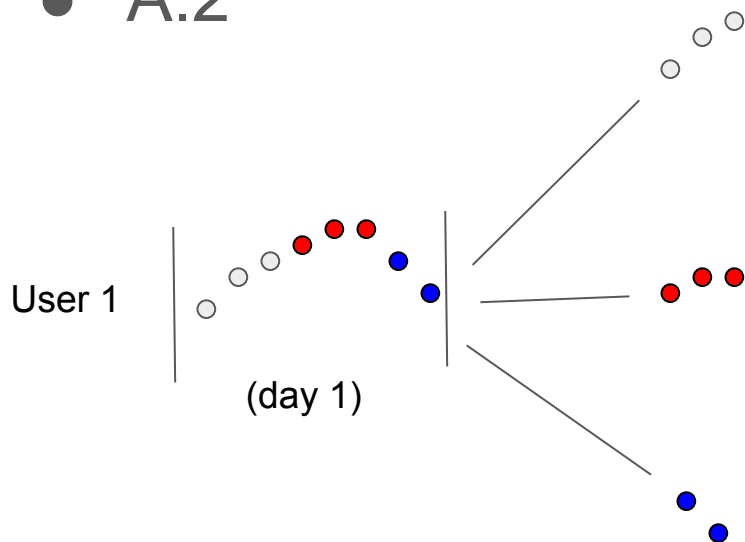# Project

● A.1

User 1

(day 1)   (day 2)   (day 3)   (day 4)   (day 5)

User 1

(day 1)

● Distance (e.g. **Haversine)**
● Speed (m/s)
● Acceleration (m/s2)
● Bearing (0 to 360 degrees)

● A.2

User 1

(day 1)

subtraj 1, class A (sample 1)
- For each: distance, speed, acceleration, and bearing
- Compute the minimum, maximum, mean, median and standard deviation. These are the trajectory feats for sample 1

subtraj 2, class B (sample 2)
- For each: distance, speed, acceleration, and bearing
- Compute the minimum, maximum, mean, median and standard deviation. These are the trajectory feats for sample 2

**subtraj 3, class C (sample 3)**
- For each: distance, speed, acceleration, and bearing
- Compute the minimum, maximum, mean, median and standard deviation. These are the trajectory feats for sample 3

Project

- ● A.3 - Explore the data.
  - ○ You are free to use any type of plots/charts.
  - ○ Answer the following question
    - ■ Is it possible to detect similarities or significant differences between the classes?
    - ■ It not a YES/NO question. Go deep into your data to find something.
    - ■ Your conclusions will suggest a hierarchy for task B.

Project

- ## B.1 - Hierarchical classification
  - Propose a structure based on findings of A.3.
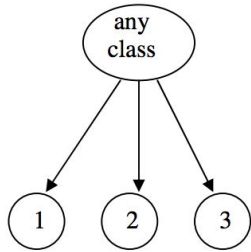
(a) Flat classification
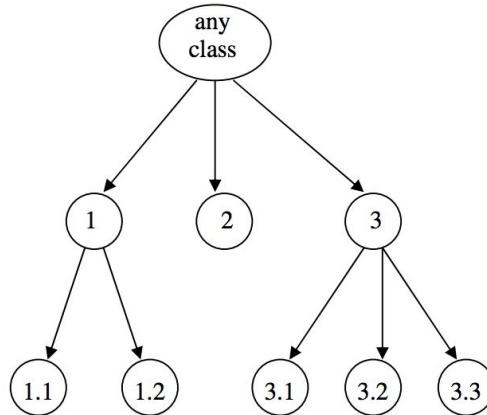
(b) Hierarchical, tree-structured classification

Organize the classes

- Bus
- Car
- Walk
- Taxi
- Subway
- Train

Project

- B.2 - **Code!**

    - We will evaluate if your proposal (B1) matches what you implemented here.
    - Try to use good coding practices.
    - **Please document your code**, you will make my life easier.

Project

- B.3 - Evaluation (**Flat** x **Hierarchical**)
  - **Choose 2 different classifiers?** Why you chose them? You are free to do anything you want in this problem (e.g. boosting, bagging, feature selection).
  - Use a stratified 10-fold cross-validation (read the scikit docs). You will evaluate the ability of the models to classify a 6-class problem. Your choice of classifiers must be guided by this multiclass problem.