

Assignment

Unsupervised Machine Learning

DATA SET: [Crowdedness at the Campus Gym](#)

Background of Data Set:

When is my university campus gym least crowded, so I know when to work out?
We measured how many people were in this gym once every 10 minutes over the last year.

TASK

Perform Principal Component Analysis and help us identify that what is the variance explained by First 4 Principal components.

DATA DESCRIPTION

The dataset consists of 26,000 people counts (about every 10 minutes) over the last year. In addition, I gathered extra info including weather and semester-specific information that might affect how crowded it is. The label is the number of people, which I'd like to predict given some subset of the features.

LABEL:

- Number of people

FEATURES:

- date (string; datetime of data)
- timestamp (int; number of seconds since beginning of day)
- day_of_week (int; 0 [monday] - 6 [sunday])
- is_weekend (int; 0 or 1) [boolean, if 1, it's either saturday or sunday, otherwise 0]
- is_holiday (int; 0 or 1) [boolean, if 1 it's a federal holiday, 0 otherwise]
- temperature (float; degrees fahrenheit)

- is_start_of_semester (int; 0 or 1) [boolean, if 1 it's the beginning of a school semester, 0 otherwise]
- month (int; 1 [jan] - 12 [dec])
- hour (int; 0 - 23)

Data set can be taken from below as well:

