Introduction into Data Science - Project Plan

Maximilian Tornow

April 19, 2018

1. What is the general topic?
   Dengue fever is a huge problem in some parts of the US, to adapt to
   the variying amounts of dengue fever patients provides a big challenge
   to state authorities.

2. What data will be used?
   Weather and environmental data of two US cities, on a monthly basis.
   Combined with the amount of dengue cases in the regarding city.

3. How will data be processed?
   This still needs to be decided in particular. But probably a estimation
   of most NaN values as means, or removal of the NaN datapoints.

4. What questions will be asked and attempted to answer?
   How many dengue fever patients will there be in regards to environ-
   mental data ?
   Which environmental datafeatures are the most important to forecast
   the projection ?

5. What analysis techniques or algorithms are planned to be used?
   The data will be analyized using mainly supervised learning algorithms
   (LDA, QDA, SVC) as well as PCA to reduce the dimension of the data
   set.