Information for assignments on electricity theft

December 16, 2016

1 Indicators

The dataset contains examples of indicators extracted from different consumers and different types of days, including an indicator of the presence of theft (there is no repetition of consumers so there is no need to take extra precautions when doing divisions between training and testing data).

For the multi-class problem: **theft type 1** indicates theft resulting in a more abrupt consumption change; **theft type 2** indicates theft resulting in more subtle pattern changes, such as in the consumption pattern shape while maintaining total consumption.

The indicators are extracted for a consumer and a certain day t using his smart metering data as represented by Figure 1.

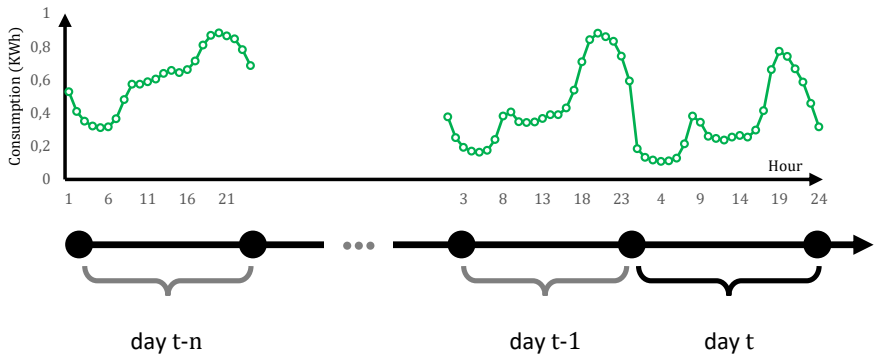

Figure 1: Example of data used for the extraction of consumption indicators.

If the theft starts on day t the change of pattern should be reflected on a change of consumption behavior in comparison with the past. If the

theft has started before it should be reflected on a change of consumption behavior in comparison to similar consumers. There are the same number of examples with theft starting at day t and theft that started before.

More information on the indicators:

- $\mathbf{I_1}$: Indicator of consumption variation. Ratio between the consumption of the last day and the last 5 days.
- $\mathbf{I_2^d}, \mathbf{I_2^c}$: Indicators of hourly consumption pattern change. Relates the hourly pattern of a day with the mean hourly pattern of the 5 days before. $\mathbf{I_2^d}$ uses the euclidean distance, changes in absolute consumption will be the most relevant for the indicator. $\mathbf{I_2^c}$ uses the Pearson correlation, changes of dynamic can be detected.
- $\mathbf{I_3}$: Indicator of consumption difference in comparison to the set of 10 consumers with the greatest similarity when comparing characteristics. Compares the mean consumption of the last 5 days to the mean consumption for the same days for the consumers with the most similar characteristics.
- $\mathbf{I_4^d}, \mathbf{I_4^c}$: Indicators of hourly consumption pattern difference in comparison to the 10 consumers with the greatest similarity when comparing characteristics. Relates the mean hourly consumption of the last 5 days between consumers. $\mathbf{I_4^d}$ uses the euclidean distance, changes in absolute consumption will be the most relevant for the indicator. $\mathbf{I_4^c}$ uses the Pearson correlation, changes of dynamic can be detected.