PERFORMANCE EVALUATION

HOMEWORK 4
LOAD PREDICTION CONTEST

Jean-Yves Le Boudec
Eleni Stai

Due Date: April 15^{th}, 2019, 23:55

1 TUTORIAL

The corresponding tutorial material for this homework is available on the homeworks web page. You are strongly advised to do the exercises from the tutorial before starting to solve the homework. But please do not assume that the forecast method in the tutorial is necessarily the one you have to use in your solution.

2 LOAD PREDICTION

The EPFL Smart Grid project (`smartgrid.epfl.ch`) was created by LCA2 and DESL labs of EPFL. As part of the project have built a monitoring infrastructure on the EPFL campus grid. It allows us to measure 50 times per second (every $20ms$) voltage and current phasors for ELA, ELE, ELG and ELL buildings of EPFL. All the measurements are available in our file repository with historical data (`http://nanotera-stg2.epfl.ch/`). We can use these files to reconstruct the behavior of the electrical network in the past. For example, we can track the energy consumption starting from measured quantities (voltages, currents).

Your goal in this exercise is to use the historical data to perform the prediction of the hourly energy consumption (aggregated for EL buildings). Such a task can be performed if you want to participate in energy market with a goal of forecasting how much energy you will need to buy for the following day.

To speed-up the things for you we give you the following on Moodle:

- .mat file that contains data points with calculated energy consumption (already aggregated). Each data point represents 1-hour energy consumption starting on October 5^{th} 2014 at 00:00:00 until May 31^{th} 2015 at 23:59:59.98. In other, words, the first 24 data points correspond to the energy consumption on October 5th etc.
- Python script (compatible with version 3.4 of Python) that you can use to fetch and process files for other days. A similar script was used to generate .mat file above. Note that data for June 2015 has been hidden.

We ask you to predict the energy consumption for three isolated hours:

- Friday June 5^{th} 2015 11h-12h
- Friday June 5^{th} 2015 21h-22h
- Saturday June 6^{th} 2015 15h-16h

Your answer should be, for each of the three hours, a 95% prediction interval $[y_{min}, y_{max}]$ for the predicted sample. So, in total, you should give three intervals. Make sure you write clearly which date corresponds to which interval.

Explain theoretically how you found the solution. Include your code. The code should be well organized and explained by comments, use suggestive names for variables.

3 THE CONTEST

In addition to the exercise, you may participate in the energy-consumption prediction contest. If you wish to participate, clearly mention so in your report. Additionally, you will need to submit your code and instructions to execute it.

We will compute your score as follows. First, we will pick three hours of our choice from June 2015. For these hours, we will execute your code to your prediction. For each prediction that your predictor gives, we interpret the answer $[y_{min}, y_{max}]$ as a 95% confidence interval of a normal distribution. We let $\mu = \frac{y_{min}+y_{max}}{2}$ and $\sigma = \frac{y_{max}-y_{min}}{3.92}$. Your score for that prediction will be computed by the Continuous Ranked Probability Score (CRPS) that is commonly used in the literature for probabilistic forecasts. CRPS is defined as follows

$$CRPS(F, x) = \int_{-\infty}^{+\infty} \left(F(y) - \mathbf{1}_{(y-x)}\right)^2 dy,$$

where x is the observation, F is the CDF associated with the empirical probabilistic forecast and $\mathbf{1}_{(y-x)}$ is the Heaviside function that is equal to 1 when $y \geq x$ else it is 0. For the computation of your score for a specific prediction, F is the normal distribution with parameters μ and σ, i.e., $\mathcal{N}(\mu, \sigma^2)$. In this case, CRPS takes a closed form:

$$CRPS(\mathcal{N}(\mu, \sigma^2), x) = \sigma \left(\frac{x-\mu}{\sigma} \left(2\Phi\left(\frac{x-\mu}{\sigma}\right) - 1 \right) + 2\phi\left(\frac{x-\mu}{\sigma}\right) - \frac{1}{\sqrt{\pi}} \right),$$

where ϕ, Φ represent the standard Gaussian probability density and cumulative distribution functions, respectively. The lower the CRPS value, the better the forecast is. Finally, the total score will be the median of the three scores.

The answer with the best score wins a dinner for two!