Prediction of Energy Consumption with Inductive Venn-Abers Predictive Distribution

Ilia Nouretdinov   I.R.NOURETDINOV@RHUL.AC.UK
James Gammerman   JGAMMERMAN@GMAIL.COM
Centre for Reliable Machine Learning, Royal Holloway, University of London, Egham, Surrey, UK

Editor: Ulf Johansson, Henrik Boström, Khuong An Nguyen, Zhiyuan Luo and Lars Carlsson

Extended Abstract

This poster presents the current state of our work in progress on development and application of Inductive Venn-Abers Predictive Distribution framework.

As a sample task, we consider real-time household energy consumption forecasting problem. Concretely, the machine learning problem is to predict evening consumption (at 18:00) based on the morning consumption (0:00–12:00) on the same day.

We use UCI public dataset on household power consumption (ECP) (1) to make prediction for the first 300 days. Only information from one particular household is used but in a fair real-time mode: in fair on-line mode: training on past days, not on the future.

The advantages of the method are following. First, outputting well-calibrated predictions that are valid in weak assumptions. Second, that this way of regression gives a rich prediction in the form of the whole predictive distribution. It can be converted to a confidence interval of any probability, with possibility of flexible choice of its location (lower or upper ray, the interval with a given centre, or of the smallest length). Third, easy adaption to non-linear dependencies.

In the previous poster (2) we also developed and demonstrated advantage of combining two ways of prediction by means of using bivariate isotonic regression as a merging tool.

The results are shown in Tab. 1. We merge two versions of k-Nearest-Neighbours algorithms with different value of k. Our evaluation criterion is Continuous Ranked Probability Score (CRPS) that is the integrated difference between CDF of the true and the predicted distributions. Like earlier in (2), the best quality is achieved in combining a relatively small (20) and large (100) values of the parameter.

The sample result for this setting is shown on Fig. 1-2. We present the predicted distributions as box-plots, the green part means $25\% - 75\%$ quantiles. The true labels are shown with black points. We present the plot in two versions: in the original chronological order, used for the on-line prediction, and re-ordered by median of the predictive distribution. In the second plot, the similar predictions are shown close to each other, so that the diversity of real labels can be visually compared to the predicted range.

Table 1: Results.

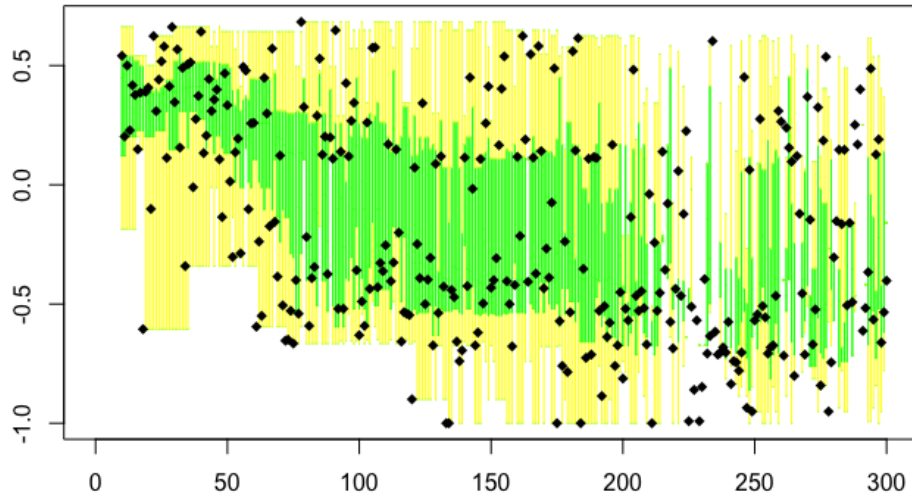| CRPS	$k_2 = 10$	$k_2 = 20$	$k_2 = 30$	$k_2 = 50$	$k_2 = 100$
$k_1 = 10$	0.6521	0.5313	0.6132	0.5738	0.5760
$k_1 = 20$		0.5364	0.5269	0.5482	**0.4994**
$k_1 = 30$			0.5813	0.5108	0.5463
$k_1 = 50$				0.5236	0.5378
$k_2 = 100$					0.5233

Figure 1: Joint box plot of the predictions. X: the day of prediction (in chronological order). Y: the energy consumption at 6pm, black for the true value, green segment for 25-75% quantile of the predicted distribution.
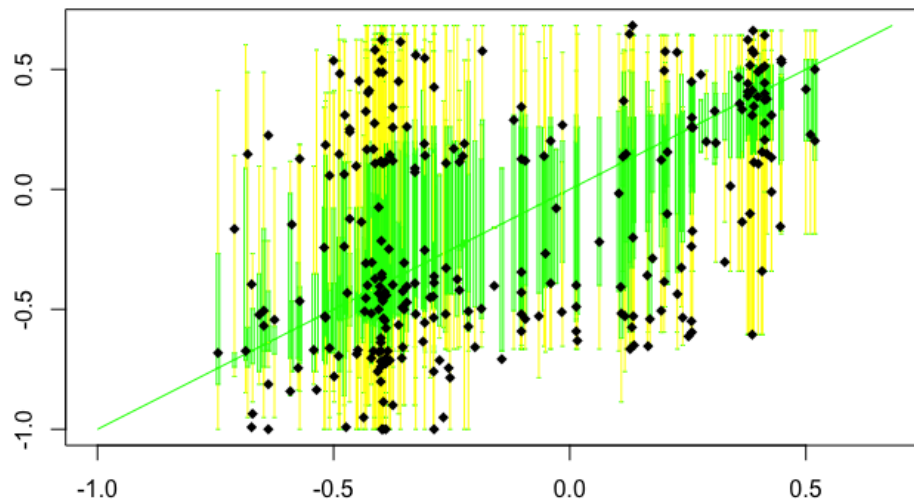

Figure 2: Re-ordered joint box plot of the predictions. X: the median of the predictive distribution. All the rest is the same.

References

[1] Individual household electric power consumption Data Set. https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption

[2] Nouretdinov, I., Gammerman, J., Rehal, D. Evaluation and Extension of Inductive Venn-Abers Predictive Distribution. Proceedings of the Ninth Symposium on Conformal and Probabilistic Prediction and Applications, 2020. https://cml.rhul.ac.uk/copa2020/posters/poster_ilia.pdf