Conformal Regressors and Predictive Systems – a Gentle Introduction

Henrik Boström

bostromh@kth.se

Division of Software and Computer Systems
Department of Computer Science
School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology

August 24, 2022

Outline

Conformal prediction

Standard and normalized split conformal regressors

Mondrian conformal regressors

Conformal predictive systems

Split conformal predictive systems

Mondrian conformal predictive systems

Can we rely on the output of our machine learning algorithms?

- Traditionally, we first generate a model and then estimate its error rate to decide on whether we are willing to rely on the model or not
- When using *conformal prediction*, we instead generate a model that we by design can rely upon; we specify a probability of error that we are willing to tolerate and the framework will guarantee that it is not exceeded

Conformal prediction

Conformal prediction turns point predictions into set predictions

- a *conformal classifier* outputs sets of class labels

  instead of e.g. $\hat{y} = \text{edible}$, a prediction may be
  $\hat{Y} = \{\text{edible}\}$, $\hat{Y} = \{\text{edible}, \text{poisonous}\}$ or even $\hat{Y} = \emptyset$

- a *conformal regressor* outputs prediction intervals

  instead of e.g. $\hat{y} = 23.5$, a prediction may be
  $\hat{Y} = [21.0, 25.0]$

Given a confidence level $1 - \epsilon$, the framework guarantees (with no stronger assumptions than the standard IID) that the probability of making an error, i.e., the correct target value is not included in the set prediction, is not larger than ϵ.

Standard conformal regressors

A *standard (inductive/split) conformal regressor* is constructed as follows:

1. randomly divide the training data into two disjoint subsets; the proper training set and the calibration set
2. train the underlying model h using the proper training set
3. calculate scores $\alpha_1, \ldots, \alpha_q$ for the calibration set, where

$$\alpha_i = |y_i - h(\boldsymbol{x}_i)|$$

4. let $\alpha_{(1)}, \ldots, \alpha_{(q)}$ be the scores sorted in descending order
5. for each test object $\boldsymbol{x}$, output the prediction interval:

$$\hat{Y}^{\epsilon} = h(\boldsymbol{x}) \pm \alpha_{(p)}$$

where $p = \lfloor \epsilon(q+1) \rfloor$ and $1 - \epsilon$ is the confidence level

Standard conformal regressors (example)

i	y_i	$h(\boldsymbol{x}_i)$	α_i	$\alpha_{(i)}$
1	25	24	1	4
2	23	24	1	3
3	27	24	3	3
4	18	22	4	2
5	12	14	2	2
6	33	34	1	1
7	19	17	2	1
8	11	14	3	1
9	14	14	0	0

$$\hat{Y}^{0.1} = h(\boldsymbol{x}) \pm \alpha_{(1)} = h(\boldsymbol{x}) \pm 4$$
$$\hat{Y}^{0.2} = h(\boldsymbol{x}) \pm \alpha_{(2)} = h(\boldsymbol{x}) \pm 3$$

Normalized conformal regressors

A *normalized (inductive/split) conformal regressor* modifies the standard conformal regressor by calculating calibration scores through:

$$\alpha_i = \frac{|y_i - h(\boldsymbol{x}_i)|}{\sigma_i}$$

where σ_i is a difficulty (quality) estimate of $\boldsymbol{x}_i$

The prediction interval at the confidence level $1 - \epsilon$ for a test instance $\boldsymbol{x}$ with difficulty σ then becomes:

$$\hat{Y}^{\epsilon} = h(\boldsymbol{x}) \pm \alpha_{(p)}\sigma$$

Normalized conformal regressors (example)

i	y_i	$h(\boldsymbol{x}_i)$	σ_i	α_i	$\alpha_{(i)}$
1	25	24	1	1	1.33
2	23	24	2	0.5	1
3	27	24	3	1	1
4	18	22	3	1.33	1
5	12	14	2	1	1
6	33	34	1	1	1
7	19	17	2	1	1
8	11	14	3	1	0.5
9	14	14	1	0	0

$$\hat{Y}^{0.1} = h(\boldsymbol{x}) \pm \alpha_{(1)}\sigma = h(\boldsymbol{x}) \pm 1.33\sigma$$
$$\hat{Y}^{0.2} = h(\boldsymbol{x}) \pm \alpha_{(2)}\sigma = h(\boldsymbol{x}) \pm \sigma$$

Notes on conformal regressors

- As the probability of error is guaranteed by construction, conformal regressors are often evaluated w.r.t. *efficiency*, i.e., the size of the prediction intervals
- The efficiency is affected by the performance of the underlying model; normalized conformal regressors are also affected by how well the difficulty estimate correlates with the actual error
- Some approaches to estimating the difficulty rely on training a separate model to predict the size of the error, e.g., using kNN or ANN; others exploit properties of the underlying model, e.g., using disagreement (variance) of the trees in a random forest

An experiment with synthetic data

A calibration set and a test set of 10 000 instances each, generated in the following way:

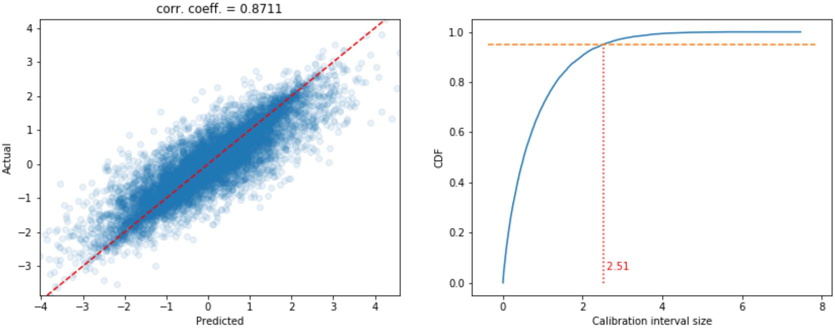
$$y_i \sim \mathcal{N}(0,1)$$

$$n_i \sim \mathcal{N}(0,1)$$

$$u_i \sim \mathcal{U}(0,1)$$

$$h(x_i) = y_i + n_i \cdot u_i$$

Synthetic data



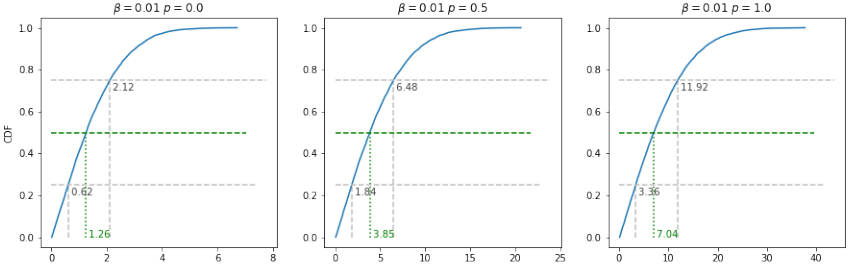
Difficulty estimates

- Difficulty estimate:

$$\sigma_i = \begin{cases} |n_i| + \beta, & \text{if } \mathbb{1}(\sim \mathcal{U}(0,1) > p) \\ | \sim \mathcal{N}(0,1)| + \beta, & \text{otherwise} \end{cases}$$

- Three levels of randomness will be considered: $p = 0$, $p = 0.5$ and $p = 1.0$
- We will set $\beta = 0.01$

Interval sizes



Observations on normalized conformal regressors

1. A less informative difficulty estimate leads to a higher variance of the interval sizes; with less information about the difficulty, the interval sizes should instead be more uniform
2. The prediction intervals can become unreasonably large; more than twice the largest observed absolute error

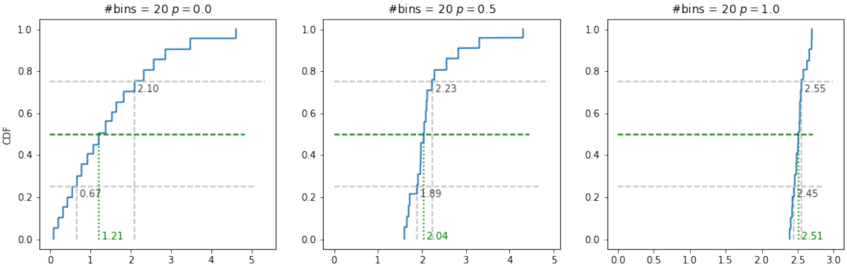
Mondrian conformal regressors

A *Mondrian conformal regressor* modifies a standard conformal regressor by dividing the calibration set into disjoint subsets according to a *Mondrian* taxonomy with k categories, and where a standard conformal regressor is produced for each subset.

A prediction interval for a test instance is obtained by assigning it to one of the k categories and using the standard conformal regressor of that category.

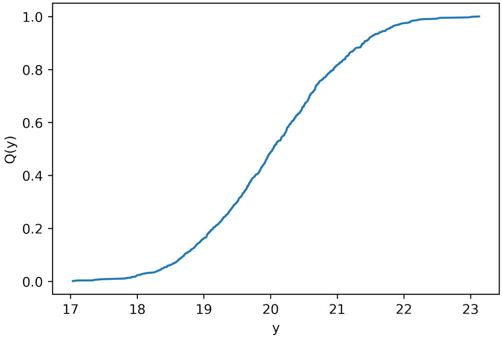
One option is to form the categories by binning of the difficulty estimates.

Interval sizes



Conformal predictive systems

Conformal predictive systems for regression output *conformal predictive distributions* (cumulative distribution functions)



Conformal predictive systems

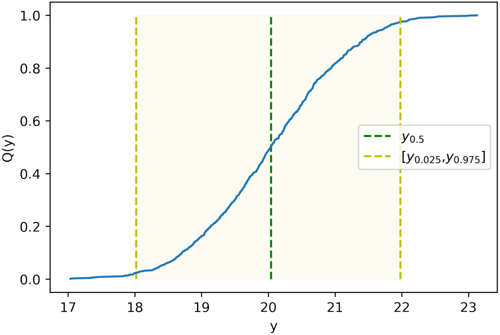
Conformal predictive systems come with a *validity* guarantee; the output of the conformal predictive distributions (the p-values) for the correct target values are distributed uniformly on $[0, 1]$.

This allows us to control the error level when making predictions on whether the correct target values are larger (or lower) than a certain threshold value, e.g., we may rule out with 99% confidence that the temperature will be above a critical level.

Conformal predictive distributions

We can also extract threshold values from the conformal predictive distributions for the p-values that we are interested in.

For example, we can obtain prediction intervals with a coverage guarantee, similar to conformal regressors.



Split conformal predictive systems

A *split conformal predictive system* can be constructed as follows:

1. randomly divide the training data into two disjoint subsets; the proper training set and the calibration set
2. train the underlying model h using the proper training set
3. calculate scores $\alpha_1, \ldots, \alpha_q$ for the calibration set, where

$$\alpha_i = \frac{y_i - h(\boldsymbol{x}_i)}{\sigma_i}$$

4. let $\alpha_{(1)}, \ldots, \alpha_{(q)}$ be the scores sorted in ascending order
5. for each test object $\boldsymbol{x}$ with difficulty σ:

   let $C_{(i)} = h(\boldsymbol{x}) + \alpha_{(i)}\sigma$ for $i \in \{1, ..., q\}$

   let $C_{(0)} = -\infty$ and $C_{(q+1)} = \infty$

   output the conformal predictive distribution:

$$Q(y) = \begin{cases} \frac{n+\tau}{q+1} & \text{if } y \in \left(C_{(n)}, C_{(n+1)}\right) \text{ for } n \in \{0, ..., q\} \\ \frac{n'-1+(n''-n'+2)\tau}{q+1} & \text{if } y = C_{(n)} \text{ for } n \in \{1, ..., q\} \end{cases}$$

Split conformal predictive systems (example)

i	y_i	$h(\boldsymbol{x}_i)$	σ_i	α_i	$\alpha_{(i)}$
1	25	24	1	1	-1.33
2	23	24	2	-0.5	-1
3	27	24	3	1	-1
4	18	22	3	-1.33	-1
5	12	14	2	-1	-0.5
6	33	34	1	-1	0
7	19	17	2	1	1
8	11	14	3	-1	1
9	14	14	1	0	1

$$C_{(1)} = h(\boldsymbol{x}) + \alpha_{(1)}\sigma, \ldots, C_{(9)} = h(\boldsymbol{x}) + \alpha_{(9)}\sigma$$

Split conformal predictive systems (example, cont.)

For $h(\boldsymbol{x}) = 20$ and $\sigma = 1$:

$$C_{(0)}, \ldots, C_{(10)} = -\infty, 18.67, 19, 19, 19, 19.5, 20, 21, 21, 21, \infty$$

$$Q(15) = \tfrac{0+\tau}{9+1} \text{ since } 15 \in \left(C_{(0)}, C_{(1)}\right)$$

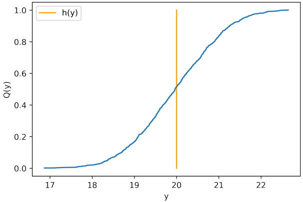
$$Q(18.8) = \tfrac{1+\tau}{9+1} \text{ since } 18.8 \in \left(C_{(1)}, C_{(2)}\right)$$

$$Q(19) = \tfrac{2-1+(4-2+2)\tau}{9+1} = \tfrac{1+4\tau}{9+1} \text{ since } 19 = C_{(2)}$$

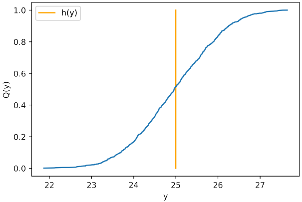
$$Q(22) = \tfrac{9+\tau}{9+1} \text{ since } 22 \in \left(C_{(9)}, C_{(10)}\right)$$

Conformal predictive distributions

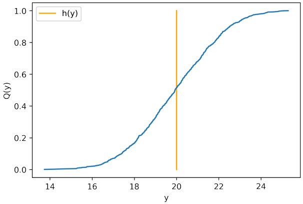
$$h(x) = 20 \ \sigma = 1$$



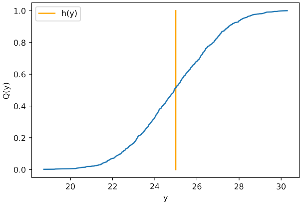
$$h(x) = 25 \ \sigma = 1$$



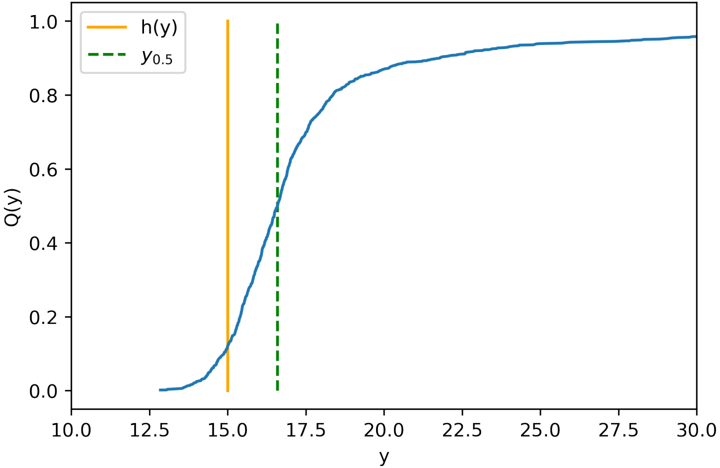
$$h(x) = 20 \ \sigma = 2$$



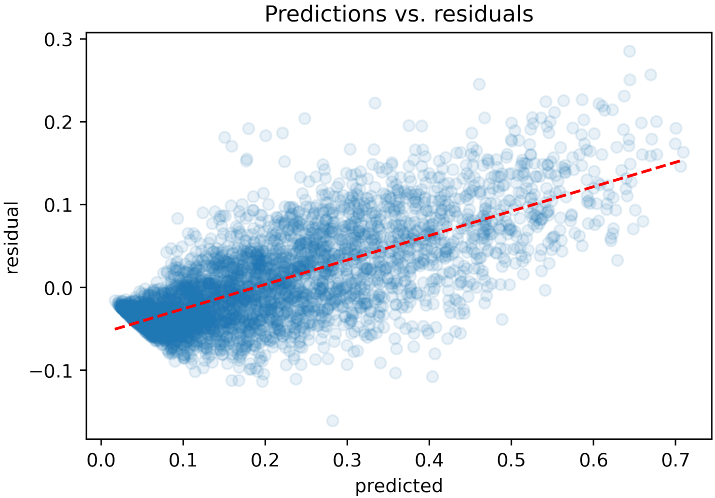
$$h(x) = 25 \ \sigma = 2$$



Calibration with conformal predictive distributions



Heteroscedastic residuals



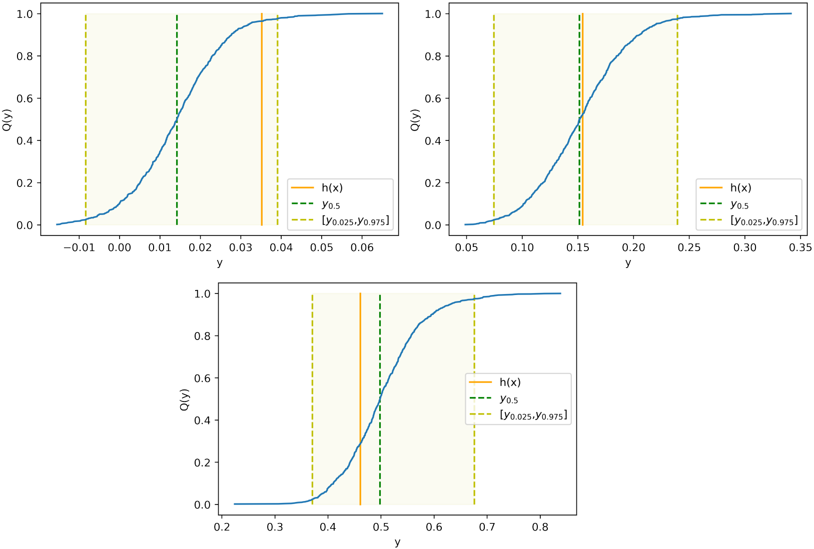
Mondrian conformal predictive systems

A *Mondrian conformal predictive system* modifies a regular conformal predictive system by dividing the calibration set into disjoint subsets according to a *Mondrian* taxonomy with k categories, and where a conformal predictive system is produced for each subset.

A conformal predictive distribution for a test instance is obtained by assigning it to one of the k categories and using the conformal predictive system of that category.

One option is to form the categories by binning of the predictions.

Random forests on the bank8fm dataset

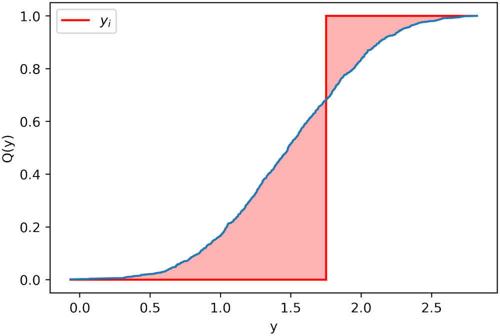


Evaluating conformal predictive systems

- The validity can be investigated by testing whether the p-values for a test set are distributed uniformly on $[0, 1]$, e.g., using the Kolmogorov-Smirnov test
- The coverage of extracted prediction intervals can also be investigated
- When using the conformal predictive distributions for calibration, the predictive performance, e.g., as measured by mean-squared error, can be compared to the original underlying model
- Continuous ranked probability score (CRPS) is another option, which uses the full conformal predictive distributions

Continuous ranked probability score (CRPS)

$$CRPS(Q, y_i) = \int_{-\infty}^{\infty} (Q(y) - \mathbb{1}(y \geq y_i))^2 dy$$



Concluding remarks

- Conformal prediction allows for controlling the probability of error by turning point predictions into set predictions
- Standard conformal regressors produce equisized intervals, while normalized conformal regressors may produce more informative and tighter intervals
- Normalized conformal regressors may however produce intervals that seem to be informative without actually being so and the sizes may be unreasonably large; a remedy for this is provided by Mondrian conformal regressors

Concluding remarks (cont.)

- Conformal predictive systems is a recent development that provides predictions in the form of conformal predictive distributions, which are more informative than prediction intervals; the latter can be derived from the former
- Mondrian conformal predictive systems allow for affecting the shape of the output conformal predictive distributions, beyond changing the location and scale; this can significantly improve performance, e.g., as measured by CRPS

Areas not covered in this tutorial

- Online (transductive) approaches
- Alternative ways of defining nonconformity scores
- Using out-of-bag predictions instead of a calibration set
- Combining multiple conformal regressors/predictive systems, e.g., multiple splits, jackknife+, cross-conformal prediction, synergy conformal prediction
- Theoretical foundations and guarantees
- Decision making using conformal regressors and predictive systems
- Adapting to scenarios in which the IID assumption is violated; see e.g. invited talk by Prof. Barber
- Practical applications, see e.g. invited talk by Dr. Carlsson and Dr. Ahlberg

Future work

- New ways of defining nonconformity scores, difficulty estimates and Mondrian categories
- New approaches to extracting intervals and point predictions from conformal predictive distributions
- New performance metrics/evaluation procedures that reflect actual use of the prediction intervals/predictive distributions

References

For theoretical foundations of conformal regressors:

- Vovk, V., Gammerman, A. and Shafer, G., 2005. Algorithmic learning in a random world. Springer.
- Vovk, V., Nouretdinov, I. and Gammerman, A., 2009. On-line predictive linear regression. The Annals of Statistics, pp.1566-1590.

Seminal paper on inductive/split conformal regressors:

- Papadopoulos, H., Proedrou, K., Vovk, V. and Gammerman, A., 2002, Inductive confidence machines for regression. In European Conference on Machine Learning, pp. 345-356.

More recent theoretical results on conformal regressors:

- Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R.J. and Wasserman, L., 2018. Distribution-free predictive inference for regression. Journal of the American Statistical Association, 113(523), pp. 1094-1111.

References (cont.)

Comparison of state-of-the-art conformal regressors and using out-of-bag predictions for calibration:

- Johansson, U., Boström, H., Löfström, T. and Linusson, H., 2014. Regression conformal prediction with random forests. Machine learning, 97(1-2), pp. 155-176.

Using variance for difficulty estimation and modified prediction procedure when using out-of-bag predictions:

- Boström, H., Linusson, H., Löfström, T. and Johansson, U., 2017. Accelerating difficulty estimation for conformal regression forests. Annals of Mathematics and Artificial Intelligence, 81(1-2), pp.125-144.

Mondrian conformal regressors:

- Boström, H. and Johansson, U., 2020. Mondrian conformal regressors. In Conformal and Probabilistic Prediction and Applications (pp. 114-133). PMLR.

References (cont.)

Seminal paper on conformal predictive systems:

- Vovk, V., Shen, J., Manokhin, V. and Xie, M.G., 2017. Nonparametric predictive distributions based on conformal prediction. In Conformal and Probabilistic Prediction and Applications (pp. 82-102). PMLR.

Split conformal predictive systems:

- Vovk, V., Petej, I., Nouretdinov, I., Manokhin, V. and Gammerman, A., 2020. Computationally efficient versions of conformal predictive distributions. Neurocomputing, 397, pp.292-308.

Mondrian conformal predictive systems:

- Boström, H., Johansson, U. and Löfström, T., 2021. Mondrian conformal predictive distributions. In Conformal and Probabilistic Prediction and Applications (pp. 24-38). PMLR.