UiO : **Department of Mathematics**
University of Oslo

An Informal Introduction to Conformal Prediction

BigInsight Day, November 21st 2022

Anders Hjort
Eiendomsverdi and University of Oslo

Introduction

- New and impressive machine learning methods emerge every day, often black box models

Introduction

- New and impressive machine learning methods emerge every day, often black box models
- For this reason, uncertainty quantification is more difficult than ever – and more important than ever!

Introduction

- New and impressive machine learning methods emerge every day, often black box models
- For this reason, uncertainty quantification is more difficult than ever – and more important than ever!
- Some methods come with a notion of uncertainty, but these are not necessarily well-calibrated and don't necessarily have theoretical guarantees

Introduction

- New and impressive machine learning methods emerge every day, often black box models
- For this reason, uncertainty quantification is more difficult than ever – and more important than ever!
- Some methods come with a notion of uncertainty, but these are not necessarily well-calibrated and don't necessarily have theoretical guarantees
- Conformal prediction (CP) returns prediction sets instead of point predictions that have theoretical guarantees regardless of underlying distribution
- Introduced by Vovk et al. 2005 and Shafer and Vovk 2007, recently renewed interest from machine learning communities

Intuition

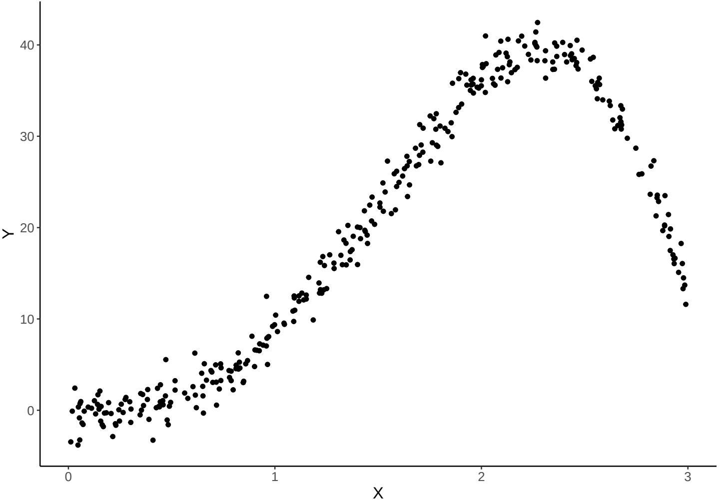


Figure: Some data (X, Y).

Intuition

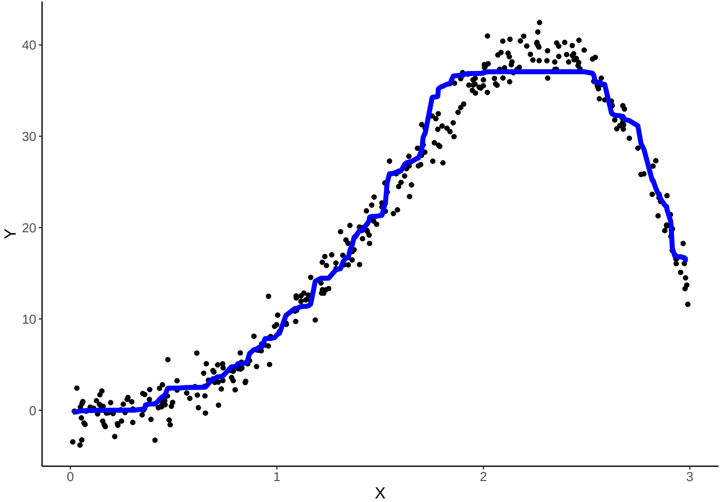


Figure: We train some black box model $f(X)$.

Intuition

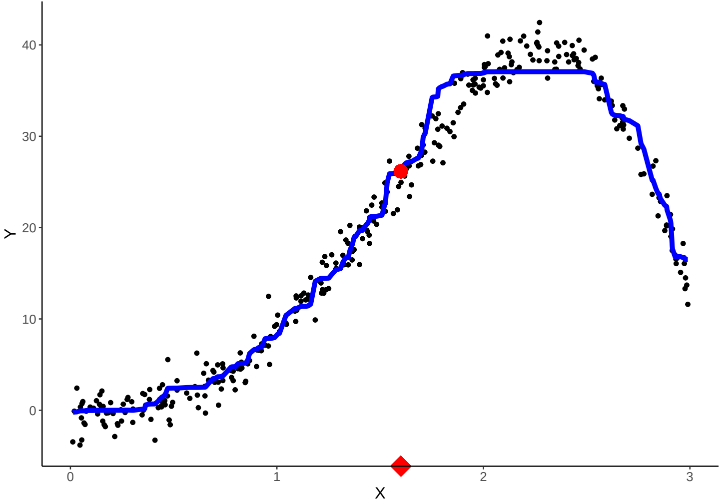


Figure: For the new point X_{new} we use the model to make a prediction.

Intuition

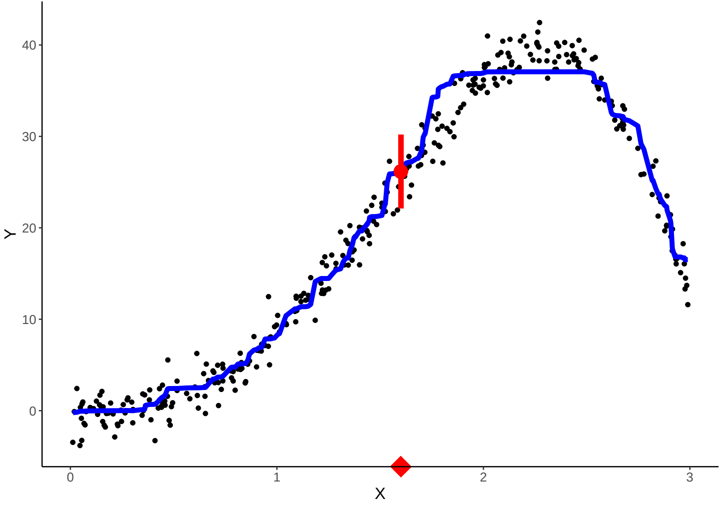
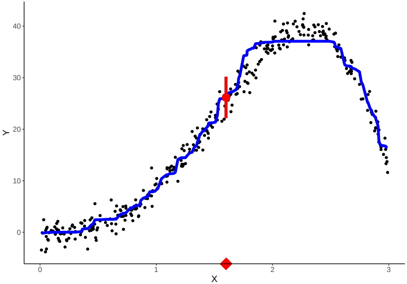


Figure: ... but how certain are we about the prediction?

Intuition



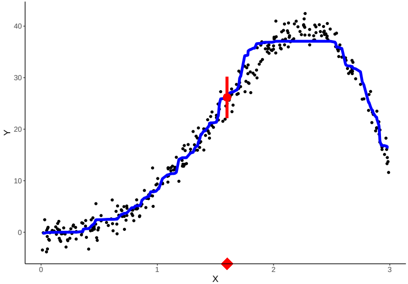
With our black box model we obtain a prediction $f(X_{new})$ for the new instance X_{new}.

In addition to the point prediction, we want to create a prediction set $C(X_{new})$ such that

$$P(Y_{new} \in C(X_{new})) \geq 1 - \alpha$$

for some $0 < \alpha < 1$.

Intuition



With our black box model we obtain a prediction $f(X_{new})$ for the new instance X_{new}.

In addition to the point prediction, we want to create a prediction set $C(X_{new})$ such that

$$P(Y_{new} \in C(X_{new})) \geq 1 - \alpha$$

for some $0 < \alpha < 1$.

In simple terms: We want to create a prediction set such that we are (e.g.) 90% sure that the true value is within the set.

Idea

- Define a score $s(X_i, Y_i)$ that quantifies how well a data point (X_i, Y_i) conforms with the rest of the data

Idea

- Define a score $s(X_i, Y_i)$ that quantifies how well a data point (X_i, Y_i) conforms with the rest of the data
- Calculate the score for every observation in a calibration set

Idea

- Define a score $s(X_i, Y_i)$ that quantifies how well a data point (X_i, Y_i) conforms with the rest of the data
- Calculate the score for every observation in a calibration set
- Use the $(1-\alpha)$th percentile of the scores on the calibration set to create prediction intervals for new, unobserved instances $(X_{new}, ?)$

Algorithm

Step 0: Prediction algorithm.

Use your favorite (black box) algorithm to obtain $f(x)$.

Algorithm

Step 0: Prediction algorithm.

Use your favorite (black box) algorithm to obtain $f(x)$.

Step 1: Non-conformity score.

A non-conformity score $s(X_i, Y_i)$ that quantifies how much (X_i, Y_i) conforms to the rest of the observations. Examples:

- $s(X_i, Y_i) = |Y_i - \bar{Y}|$
- $s(X_i, Y_i) = |Y_i - f(X_i)|$

Algorithm

Step 0: Prediction algorithm.

Use your favorite (black box) algorithm to obtain $f(x)$.

Step 1: Non-conformity score.

A non-conformity score $s(X_i, Y_i)$ that quantifies how much (X_i, Y_i) conforms to the rest of the observations. Examples:

- $s(X_i, Y_i) = |Y_i - \bar{Y}|$
- $s(X_i, Y_i) = |Y_i - f(X_i)|$

Step 2: Calculate non-conformity scores on calibration set.

Calculate $s(X_i, Y_i)$ for every observation in a calibration set.

Algorithm

Step 0: Prediction algorithm.

Use your favorite (black box) algorithm to obtain $f(x)$.

Step 1: Non-conformity score.

A non-conformity score $s(X_i, Y_i)$ that quantifies how much (X_i, Y_i) conforms to the rest of the observations. Examples:

- $s(X_i, Y_i) = |Y_i - \bar{Y}|$
- $s(X_i, Y_i) = |Y_i - f(X_i)|$

Step 2: Calculate non-conformity scores on calibration set.

Calculate $s(X_i, Y_i)$ for every observation in a calibration set.

Step 3: Find the correct threshold.

Let q_{90} be the 90th percentile of $s(X_1, Y_1), ..., s(X_N, Y_N)$.

Algorithm

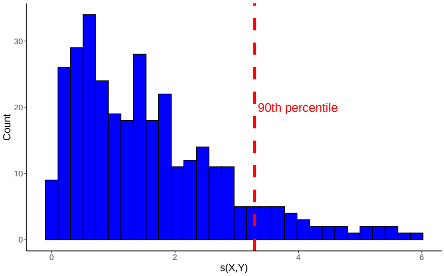
Step 0: Prediction algorithm.
Use your favorite (black box) algorithm to obtain $f(x)$.
Step 1: Non-conformity score.
A non-conformity score $s(X_i, Y_i)$ that quantifies how much (X_i, Y_i) conforms to the rest of the observations. Examples:

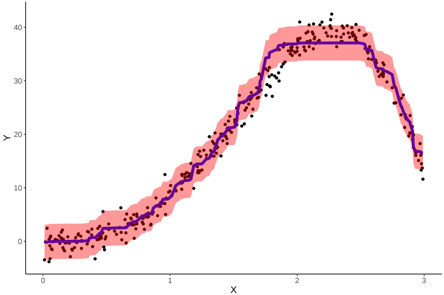
- $s(X_i, Y_i) = |Y_i - \bar{Y}|$
- $s(X_i, Y_i) = |Y_i - f(X_i)|$

Step 2: Calculate non-conformity scores on calibration set.
Calculate $s(X_i, Y_i)$ for every observation in a calibration set.
Step 3: Find the correct threshold.
Let q_{90} be the 90th percentile of $s(X_1, Y_1), ..., s(X_N, Y_N)$.
Step 4: Prediction. For a new observation $(X_{new}, ?)$ from the test set, the prediction set is

$$C(X_{n+1}) = [f(X_{new}) - q_{90}, f(X_{new}) + q_{90}]$$

A Simple Example





The 90th percentile of $s(X_i, Y_i) = |Y_i - f(X_i)|$ on a calibration is ≈ 3.3, so the prediction set is $C(X_{new}) = [f(X_{new}) - 3.3, f(X_{new}) + 3.3]$.

Assumptions and proofs

- Prediction sets give finite sample coverage guarantee:

$$P(Y_{new} \in C(X_{new})) \geq 1 - \alpha,$$

for any α, as long as (X_{new}, Y_{new}) is exchangeable with training and calibration data.

Assumptions and proofs

- Prediction sets give finite sample coverage guarantee:

$$P(Y_{new} \in C(X_{new})) \geq 1 - \alpha,$$

  for any α, as long as (X_{new}, Y_{new}) is exchangeable with training and calibration data.
- The rank of $s(X_{new}, Y_{new})$ is uniformly distributed among the previous $s(X_1, Y_1), ..., s(X_N, Y_N)$ as long as they are exchangeable!

Example with heteroscedastic errors

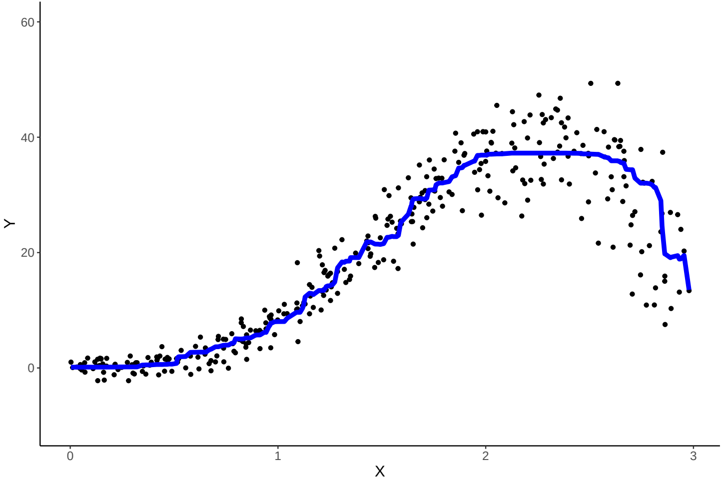


Figure: What if training data has heteroscedastic errors?

Example with heteroscedastic errors

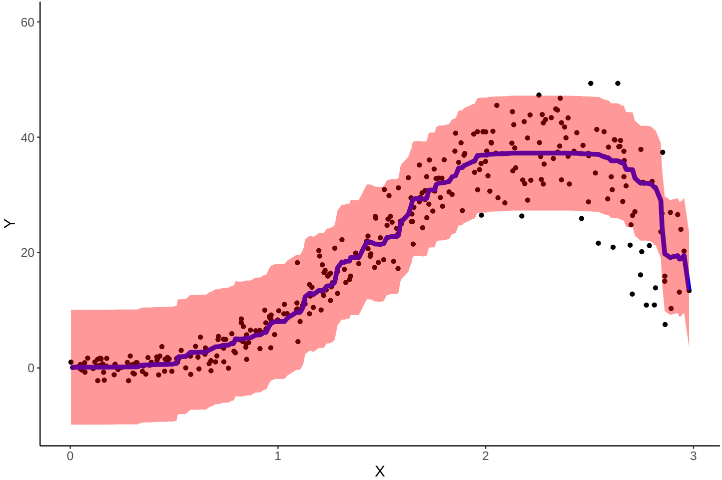


Figure: Still 90% coverage, but not so helpful.

Any non-conformity score is valid!

Interestingly, any choice of non-conformity score $s(X, Y)$ gives valid prediction sets! A common choice to handle heteroscedasticity:

$$s(X_i, Y_i) = \frac{|Y_i - f(X_i)|}{\hat{\sigma}(X_i)},$$

where $\hat{\sigma}(X_i)$ is an estimate of the standard deviation of the errors.

Any non-conformity score is valid!

Interestingly, any choice of non-conformity score $s(X, Y)$ gives valid prediction sets! A common choice to handle heteroscedasticity:

$$s(X_i, Y_i) = \frac{|Y_i - f(X_i)|}{\hat{\sigma}(X_i)},$$

where $\hat{\sigma}(X_i)$ is an estimate of the standard deviation of the errors. Instead of using

$$C(X_{n+1}) = [f(X_{new}) \pm q_{90}],$$

we can use

$$C(X_{n+1}) = [f(X_{new}) \pm \hat{\sigma}(X_{new}) \cdot q_{90}],$$

to create adaptive intervals.

Example with heteroscedastic errors

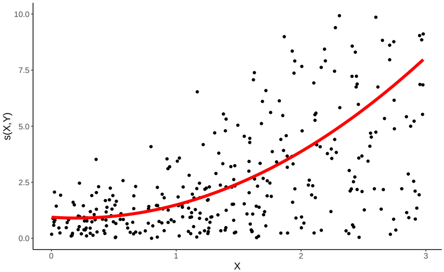


Figure: A function $\hat{\sigma}(X)$ to estimate the heteroscedasticity.

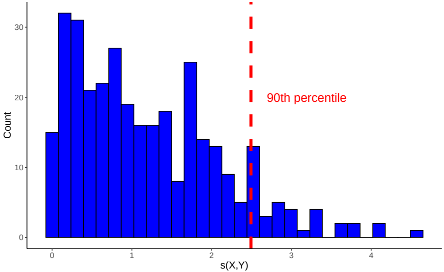


Figure: A histogram of the normalized residuals $s(X_i, Y_i) = \frac{|Y_i - f(X_i)|}{\hat{\sigma}(X_i)}$.

Example with heteroscedastic errors

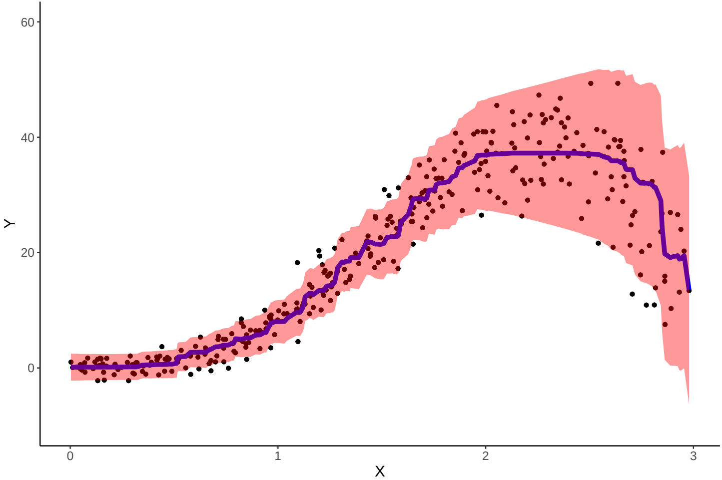


Figure: Adaptive intervals!

Conclusion

- CP can create prediction sets around your favorite point prediction method.
- CP gives coverage guarantees as long as we have exchangeable data.
- Garbage in, garbage out: Everything depends on a good point prediction and the choice of non-conformity score.
- Exciting topic with lots of potential for researchers and practitioners!

Conclusion

- CP can create prediction sets around your favorite point prediction method.
- CP gives coverage guarantees as long as we have exchangeable data.
- Garbage in, garbage out: Everything depends on a good point prediction and the choice of non-conformity score.
- Exciting topic with lots of potential for researchers and practitioners!
- Recent trends:
  - Conformal prediction beyond exchangeability (Barber et al. 2022)
  - Conformalizing other methods, such as Conformalized Quantile Regression (Romano et al. 2019)
  - Conformal predictive distributions (Vovk et al. 2017)
  - Creating tailored non-conformity scores for specific applications

References I

- Barber, Rina Foygel, Emmanuel J. Candes, Aaditya Ramdas and Ryan J. Tibshirani (2022). *Conformal prediction beyond exchangeability*. URL: https://arxiv.org/abs/2202.13415.
- Romano, Yaniv, Evan Patterson and Emmanuel Candes (2019). 'Conformalized Quantile Regression'. In: *Advances in Neural Information Processing Systems*. Vol. 32.
- Shafer, Glenn and Vladimir Vovk (July 2007). 'A tutorial on conformal prediction'. In: *Journal of Machine Learning Research* 9.
- Vovk, Vladimir, Alex Gammerman and Glenn Shafer (2005). *Algorithmic Learning in a Random World*. Springer-Verlag, Heidelberg.
- Vovk, Vladimir, Jieli Shen, Valery Manokhin and Min-ge Xie (2017). 'Nonparametric predictive distributions based on conformal prediction'. In: *Proceedings of the Sixth Workshop on Conformal and Probabilistic Prediction and Applications*. Vol. 60, pp. 82–102.

Application: House price prediction

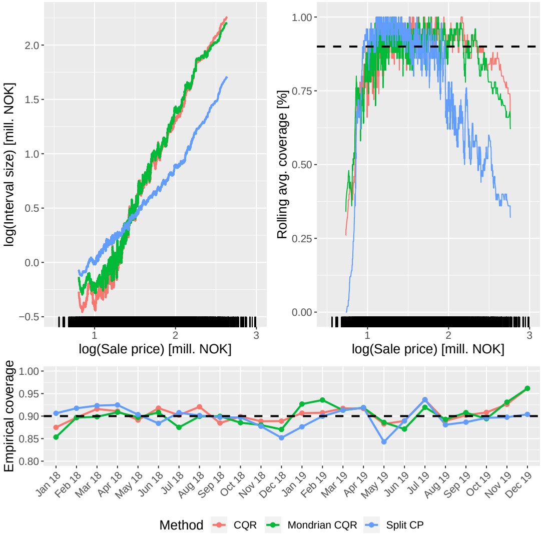
- Goal: Predict house price (Y) of a house given coordinates, size (m^2), number of bedrooms, neighborhood characteristics, etc.
- $N \approx 30\,000$ from Oslo, 2018-2019 (train/test/calibration: 1/3 each)
- Point prediction: Random forest with 500 trees
- Three CP methods for uncertainty quantification:
  - Split Conformal Prediction
  - Conformalized Quantile Regression (CQR)
  - Mondrian CQR
- Research in collaboration with Eiendomsverdi AS, presented at COPA 2022

Application: House price prediction

Table: Results from the Oslo data set at confidence level $\alpha = 0.1$. Interval sizes are given in million NOK.

Method	Coverage (%)	Mean interval size	Median interval size
Split CP	89.54	1.85	1.61
CQR	90.25	1.79	1.23
Mondrian CQR	90.40	1.85	1.25

Application: House price prediction



UiO : Department of Mathematics
University of Oslo

Anders Hjort
Eiendomsverdi and University of Oslo

An Informal Introduction to Conformal Prediction

BigInsight Day, November 21st 2022