

Conformalized Quantile Regression

Emmanuel Candès



Aad van der Vaart @60, Leiden, June 2019

Collaborators



Yaniv Romano



Evan Patterson

Thanks to...



Rina F. Barber



Aadita Ramdas



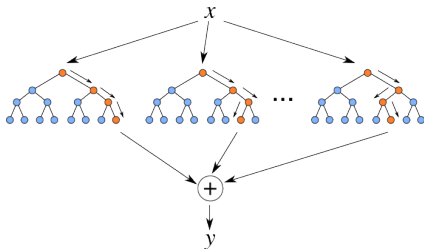
Ryan Tibshirani

BaCaRaTi papers

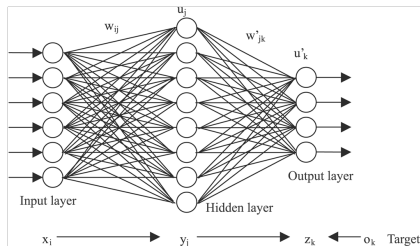
- The limits of distribution-free conditional predictive inference ('19)
- Conformal prediction under covariate shift ('19)
- The Jackknife+ ('19)

Today's predictive algorithms

random forests, gradient boosting



neural networks



Tantalizing performance in some applications

Reliability in mission critical systems

ML 15 years ago: predict movie ratings

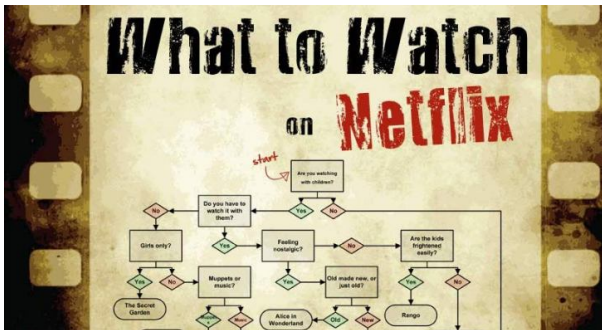


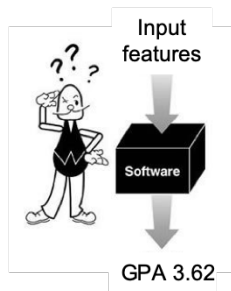
Image credit: Silveroak Casino

Reliability in mission critical systems

ML 15 years ago: predict movie ratings

ML today: who gets into college? who gets paroled, bailed? ...

Can we trust this?
 $3.62 \pm ?$



Desperately need reliable systems

Still want to use powerful, complex algorithms

Prediction intervals

Training data $(X_1, Y_1), \dots, (X_n, Y_n)$ and test point $(X_{n+1}, ?)$
(assumed exchangeable, e.g. i.i.d. from P_{XY})

$X_i \in \mathbb{R}^p$: features (e.g. high school identifier, high-school GPA, SAT score...)

$Y_i \in \mathbb{R}$: response (e.g. college GPA)

Goal: construct **marginal distribution free prediction interval**

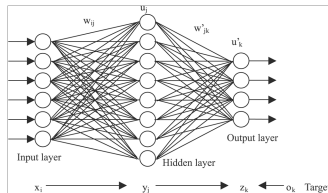
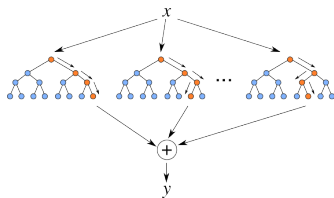
$$\mathbb{P}\{Y_{n+1} \in C(X_{n+1})\} \geq 1 - \alpha$$

- Any dist. P_{XY} (assumed unknown)
- Any sample size n

“Based on the candidate’s, high school identifier, high-school GPA, SAT score, and other attributes, the college GPA is predicted to fall in the $[3.4, 3.8]$ range”

Statistical efficiency

- Better predictions (via black box) $\rightsquigarrow$ shorter intervals!



“Candidate’s expected GPA is in range [3.4, 3.8]”

Informative

Vs.

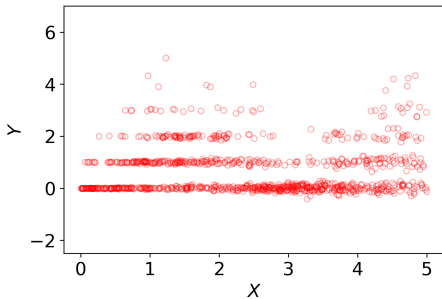
“Candidate’s expected GPA is in range [1.1, 4.3]”

Uninformative

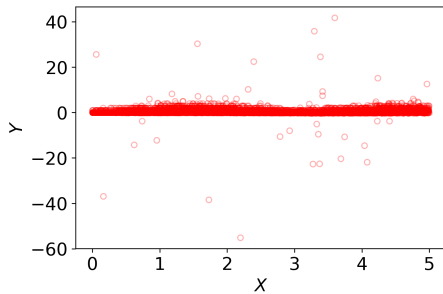
- **Heteroscedastic** data requires adjusting lengths of intervals
 $\rightsquigarrow$ to get shorter ranges

Heteroscedastic data with outliers

zoom-in

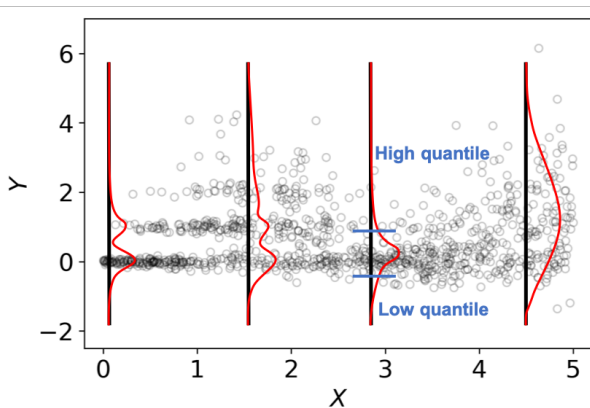


full-range view



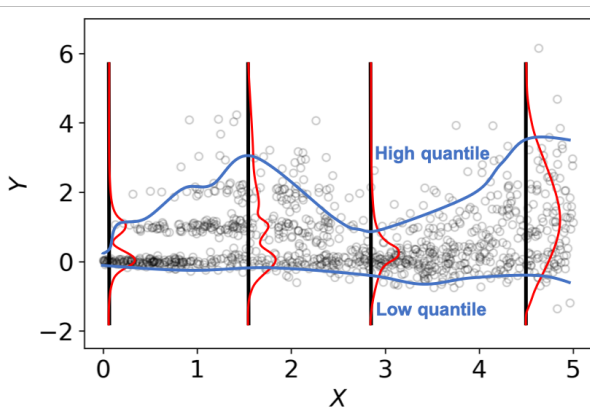
Setting with perfect knowledge

$P_{Y|X}$ known $\rightsquigarrow$ can fit upper and lower quantile functions



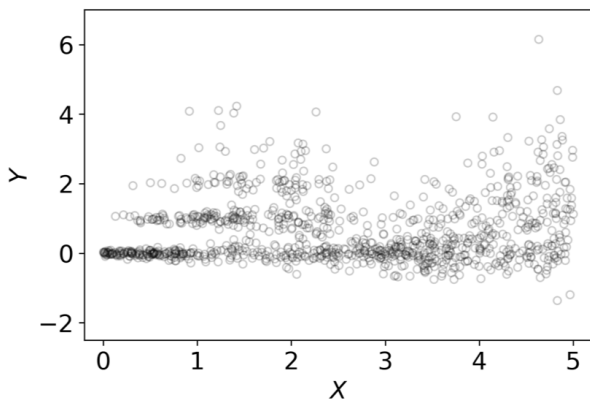
Setting with perfect knowledge

$P_{Y|X}$ known $\rightsquigarrow$ can fit upper and lower quantile functions



Length of interval can greatly vary

No perfect knowledge, only a few samples from $P_{Y|X}$!



Econometrica, Vol. 46, No. 1 (January, 1978)

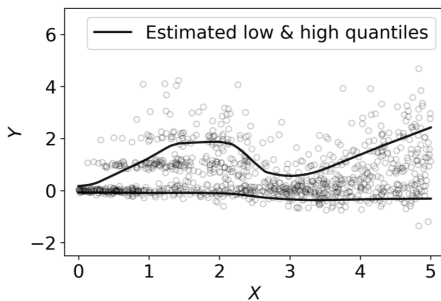
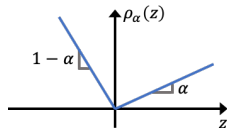
REGRESSION QUANTILES¹

BY ROGER KOENKER AND GILBERT BASSETT, JR.

Formulate quantile estimation as a learning task

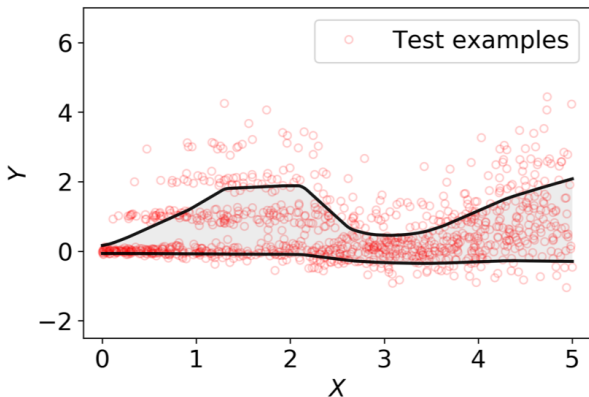
$$f(\cdot) = \operatorname{argmin}_{f \in \mathcal{F}} \sum_i \rho_\alpha(Y_i - f(X_i)) + \mathcal{R}(f)$$

- $\mathcal{R}(f)$ is a possible regularizer
- ρ_α is pinball loss Koenker & Bassett '78



Validity for unseen data?

Valid? No (imagine training a neural net)



Target coverage level: 90%; Actual coverage: 66.77%

Validity for unseen data?

- No finite sample guarantees
- Consistency ($n \rightarrow \infty$) results?
only under regularity conditions and specific models

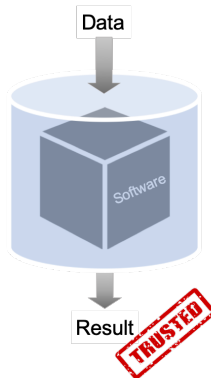
Takeuchi et al. '06, Meinshausen '06, Steinwart & Christmann '11

This talk: quantile regression with finite sample validity

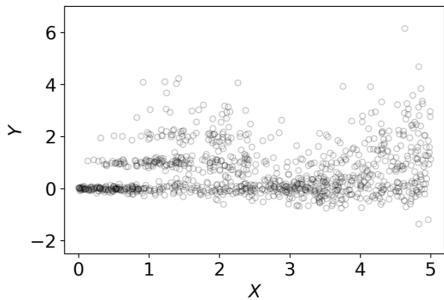
Conformalized Quantile Regression (CQR)

Wrapper around any predictive algorithm (neural nets, random forests, ...)
with perfect coverage guarantees

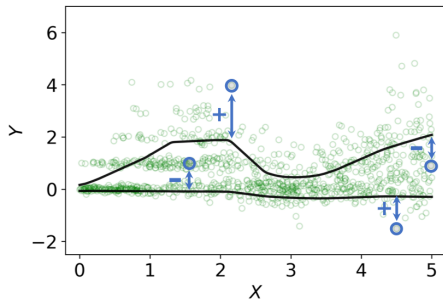
- ✓ Under finite sample
- ✓ Any dimension
- ✓ Distribution (or model) free
- ✓ Any black box



Split

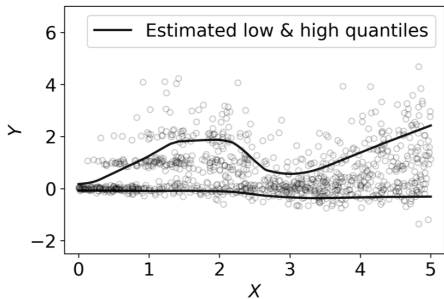


Proper training set

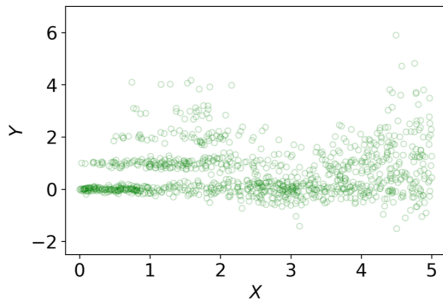


Calibration set

Fit

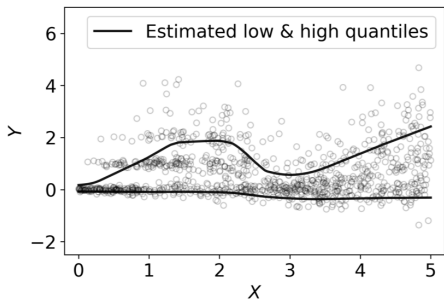


Apply quantile regression

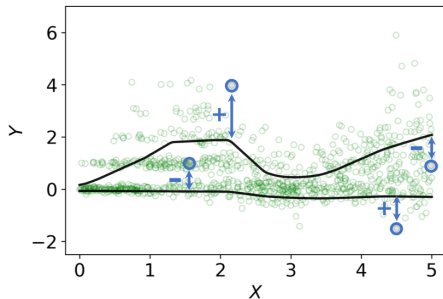


Calibration set

Calibration



Apply quantile regression



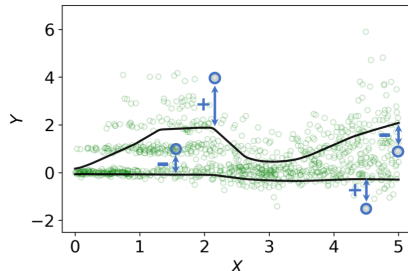
Calibrate

Calibration: how?

$$E_i = \max\{\text{lower}(X_i) - Y_i, Y_i - \text{upper}(X_i)\}$$
$$i = 1, \dots, m$$

signed distance to boundary

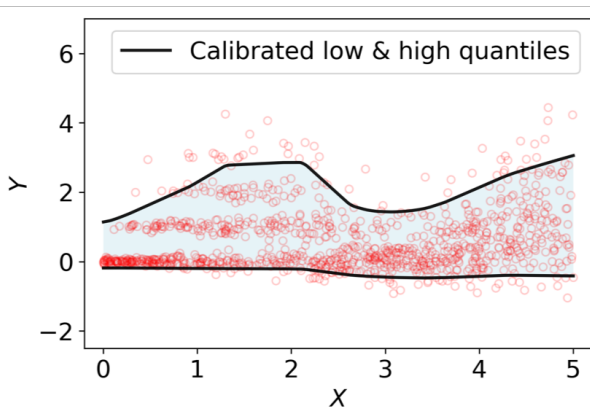
Q is $(1 - \alpha)$ th quantile of the E_i 's



Prediction interval:

$$C(X_{n+1}) = [\text{lower}(X_{n+1}) - Q, \text{upper}(X_{n+1}) + Q]$$

Validity on **new** data



Target coverage level: 90%; Actual coverage: 90.09%

Correctness of predicted range

Theorem

If (X_i, Y_i) , $i = 1, \dots, n + 1$ are exchangeable, then

$$1 - \alpha \leq \mathbb{P}\{Y_{n+1} \in C(X_{n+1})\} \leq 1 - \alpha + 1/m$$

- m is size of calibration set
- Upper bound holds if scores E_i are distinct

- ✓ Any distribution P_{XY}
- ✓ Any sample size
- ✓ Regardless of choice or accuracy of quantile regression estimate

Related work

Inspiration: conformal inference

Vovk et al. '99, Papadopoulos et al. '12



Vladimir Vovk

Classic conformal prediction Vovk et al. '99, Papadopoulos et al. '12, Lei et al. '18

- Fit **classic regression** function on proper training set

$$\mu(\cdot) = \operatorname{argmin}_{\mu \in \mathcal{F}} \sum_i (Y_i - \mu(X_i))^2 + \mathcal{R}(\mu)$$

$\mathcal{R}$
possible regularizer

- Calibration using

$$R_i = |Y_i - \operatorname{mean}(X_i)| \quad i = 1, \dots, m$$

Q is $(1 - \alpha)$ th quantile of R_i

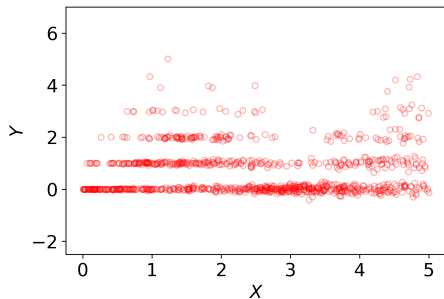
Prediction interval:

$$C(X_{n+1}) = \operatorname{mean}(X_{n+1}) \pm Q$$

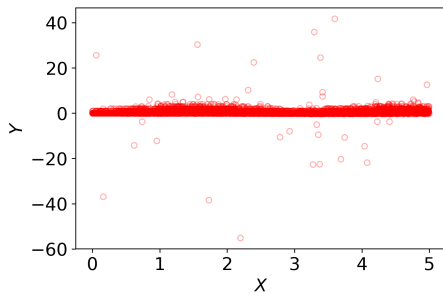
- **Major limitation: width is fixed**, equal to $2Q$ (indep. of query point)

Heteroscedastic data with outliers

Test data (zoom-in)



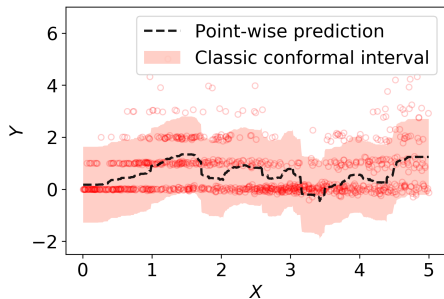
Test data (full-range view)



Comparison to split conformal prediction

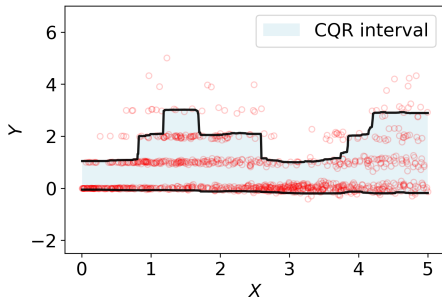
Papadopoulos et al. '12

Split conformal



Avg. Coverage 91.40%
Avg. Length 2.91

CQR



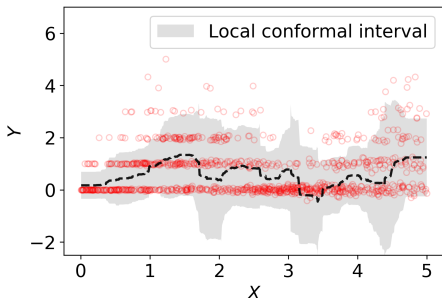
Avg. Coverage 91.06%
Avg. Length 1.99

CQR is adaptive while split conformal is not

Comparison to local conformal

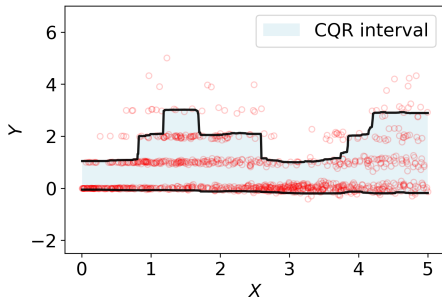
Papadopoulos et al. '08, Lei et al. '18

Local conformal (mean+variance)



Avg. Coverage 91.70%
Avg. Length 2.86

CQR



Avg. Coverage 91.06%
Avg. Length 1.99

Locally weighted conformal

Papadopoulos et al. '08, Lei et al. '18

To handle heteroscedasticity, work with relative residuals

$$\tilde{R}_i = \frac{|Y_i - \text{mean}(X_i)|}{\hat{\sigma}(X_i)} = \frac{R_i}{\hat{\sigma}(X_i)}$$

$\hat{\sigma}(X_i)$ is measure of residual dispersion; e.g., estimate the conditional MAD of $|Y_i - \text{mean}(x)|$ at $X_i = x$

Prediction interval:

$$C(X_{n+1}) = \text{mean}(X_{n+1}) \pm Q \hat{\sigma}(X_{n+1}) \quad Q \text{ is } (1 - \alpha)\text{th quantile of } \tilde{R}_i$$

Major limitation: $\hat{\sigma}$ underestimates prediction errors as it must fit to training residuals

In NNets, usual overfitting so that proper training residuals ≈ 0

$\implies$ forces Q to be large

$\implies$ intervals are less adaptive

Predicting utilization of medical services

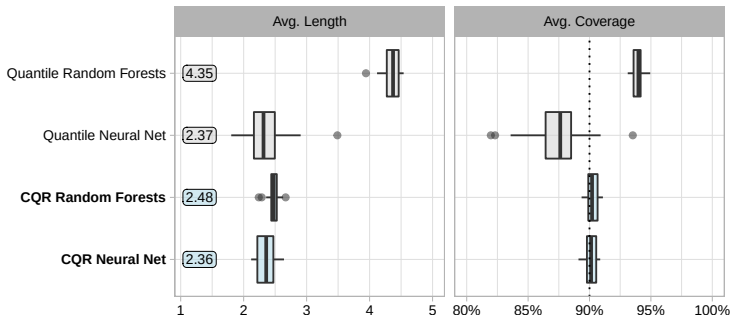
Medical Expenditure Panel Survey 2015

- X_i – age, marital status, race, poverty status, functional limitations, health status, health insurance type, ...
- Y_i – health care system utilization, reflecting # visits to doctor's office/hospital, ...
- $\approx 16,000$ subjects
- ≈ 140 features



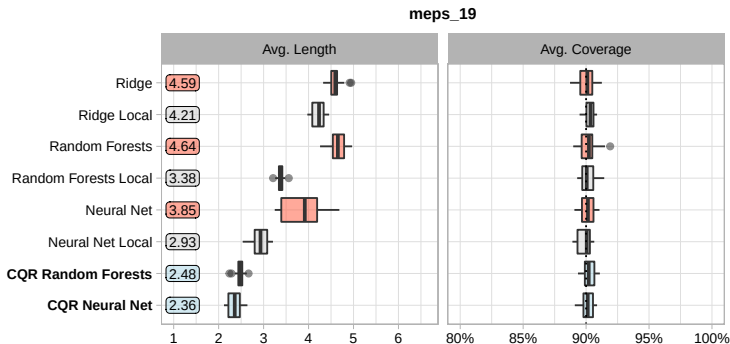
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care

Real data experiment

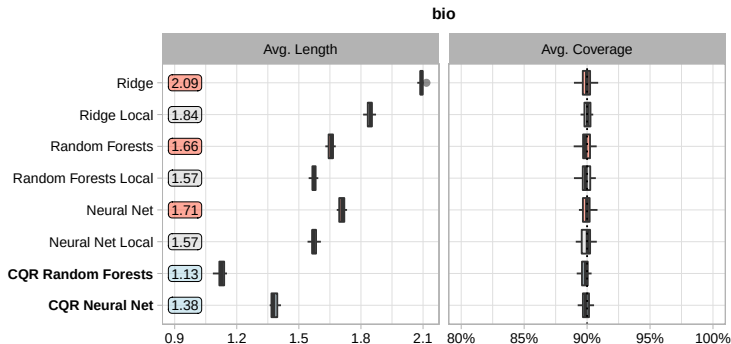


- CQR achieves exact coverage while black-box does not!
- Competitive intervals (even shorter)

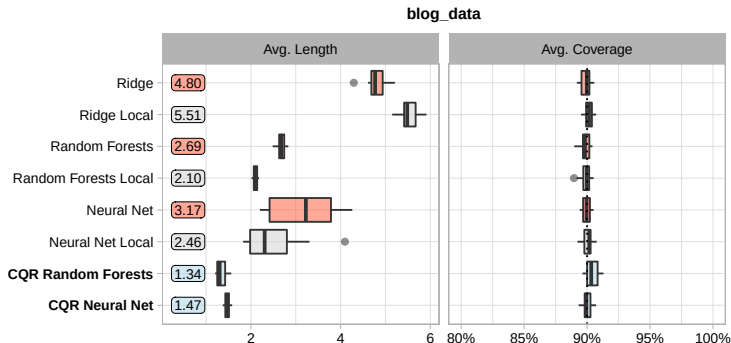
Comparison to classic & local conformal: MEPS



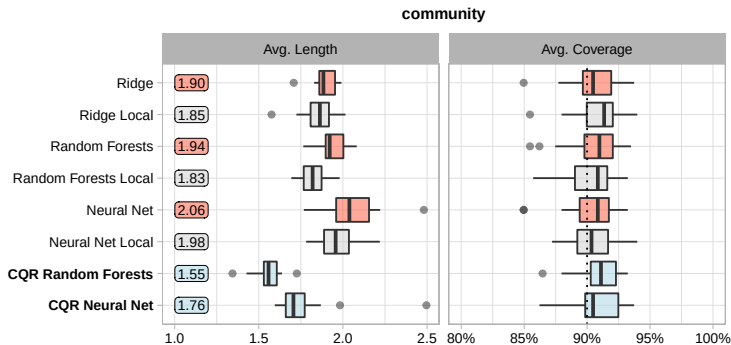
Comparison to classic & local conformal: bio



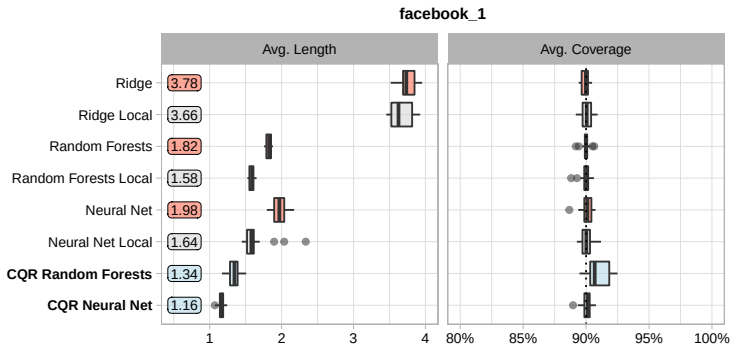
Comparison to classic & local conformal: blog data



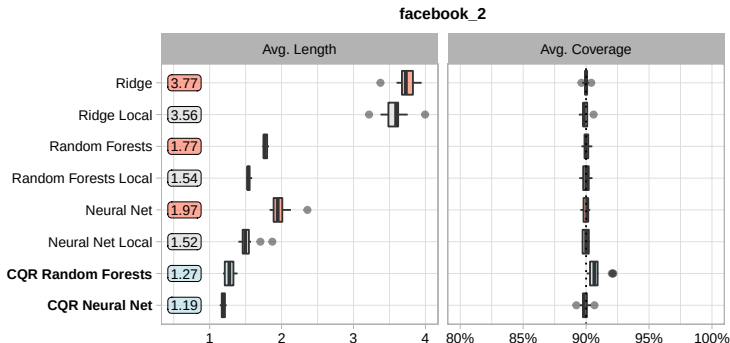
Comparison to classic & local conformal: community



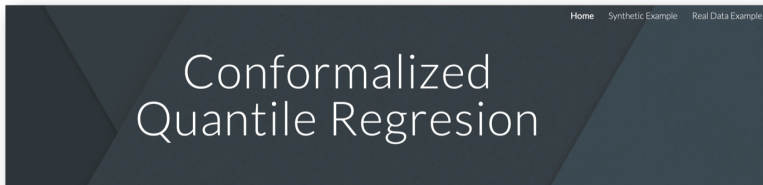
Comparison to classic & local conformal: facebook (ver. 1)



Comparison to classic & local conformal: facebook (ver. 2)



CQR: website & code



Overview

[Conformalized quantile regression](#) (CQR) is a technique for constructing prediction intervals assumptions. It combines the statistical efficiency of quantile regression with the distribution

For more information, please read the accompanying [paper](#):

Yaniv Romano, Evan Patterson, and Emmanuel J. Candes, "Confor

Code

The methods described in the paper are implemented in Python, and available under the MIT

The code is self-contained and publicly available from GitHub: <https://github.com/yromano/cqr>

Some of the code is taken from the nonconformist package available at <https://pithub.com/dlo> applications of conformal prediction.

Please contact [Yaniv Romano](#) for bug reports. Email: yaniv dot romano at gmail dot com

We now turn to invoke the CQR procedure. The class `AllQNetRegressorAdapter` defines the underlying neural network estimator. Just as before, `RegressorNc` defines the CQR object, which uses `QuantileRegErrFunc` as the nonconformity score. The function `run_icp` returns the conformal band, computed on test data. Lastly, we compute the average coverage and length using `compute_coverage`.

```
In [7]: # define quantile neural network model
quantile_estimator = helper.AllQNetRegressorAdapter(model=None,
                                                    fit_params=None,
                                                    in_shape=in_shape,
                                                    hidden_size=hidden_size,
                                                    quantiles=quantiles_net,
                                                    learn_func=nn_learn_func,
                                                    epochs=epochs,
                                                    batch_size=batch_size,
                                                    dropout=dropout,
                                                    lr=lr,
                                                    wd=wd,
                                                    test_ratio=cv_test_ratio,
                                                    random_state=cv_random_state,
                                                    use_rearrangement=False)

# define the CQR object, computing the absolute residual error of points
# located outside the estimated quantile neural network band
nc = RegressorNc(quantile_estimator, QuantileRegErrFunc())

# run CQR procedure
y_lower, y_upper = helper.run_icp(nc, x_train, y_train, x_test, idx_train, idx_cal, alpha)

# compute and print average coverage and average length
coverage_cp_qnet, length_cp_qnet = helper.compute_coverage(y_test,
                                                            y_lower,
                                                            y_upper,
                                                            x,
                                                            alpha,
                                                            "CQR Neural Net")
```

CQR Neural Net: Percentage in the range (expecting 90.00): 90.977444

CQR Neural Net: Average length: 1.671457

Summary

New prediction intervals using quantile regression

- If exchangeable observations, rigorous type-1 error in finite samples
- Outperforms existing conformal methods (in our experiments, **10/11 regression datasets**)

Y. Romano, E. Patterson, and E. J. Candès. “Conformalized quantile regression”. 2019.

<http://statweb.stanford.edu/~candes/papers/CQR.pdf>

Thank you and Happy Birthday, Aad!



Experiments

Setup

- 11 popular regression datasets
- 10 methods
 - black box – ridge, random forests, neural net
 - (3) classic conformal
 - (3) local conformal
 - (2) CQR
 - (2) *non-conformalized quantile regression
- 20 random train/test splits (80% / 20%)
- 10% desired coverage rate
- 2,200 experiments in total

Summary

Method	Avg. Length	Avg. Coverage
Ridge	3.06	90.03
Ridge Local	2.94	90.13
Random Forests	2.24	89.99
Random Forests Local	1.82	89.95
Neural Net	2.16	89.92
Neural Net Local	1.81	89.95
CQR Random Forests	1.41	90.33
CQR Neural Net	1.40	90.05
*Quantile Random Forests	*2.23	*92.62
*Quantile Neural Net	*1.49	*88.51

- Local conformal is better than classic approach
- CQR outperforms all methods, including the non-conformalized ones
- *Quantile Random Forest is too conservative
 - *Quantile Neural Net does not attain valid coverage