Conformal Prediction via Regression-As-Classification

Etash Kumar Guha[1,4], Shlok Natarajan[2], Thomas Möllenhoff[1],
Mohammad Emtiyaz Khan[1], Eugene Ndiaye[3]

[1] RIKEN AIP [2] Salesforce [3] Apple [4] SambaNova Systems

NEURAL INFORMATION PROCESSING SYSTEMS

Overview

Goal: *Generate a prediction set that guarantees coverage and is small given data*

$$\mathcal{D}_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}.$$

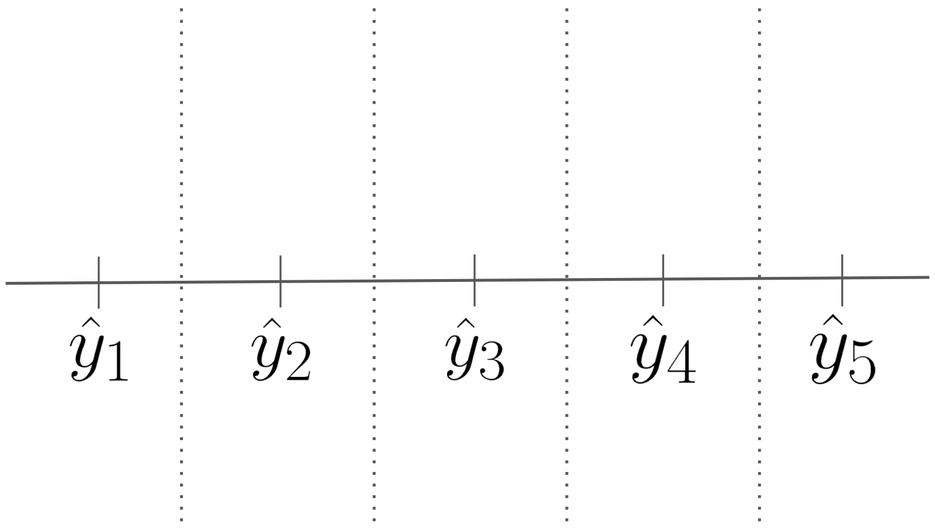
Conformal Prediction can use conditional label density

$$P(y|x)$$

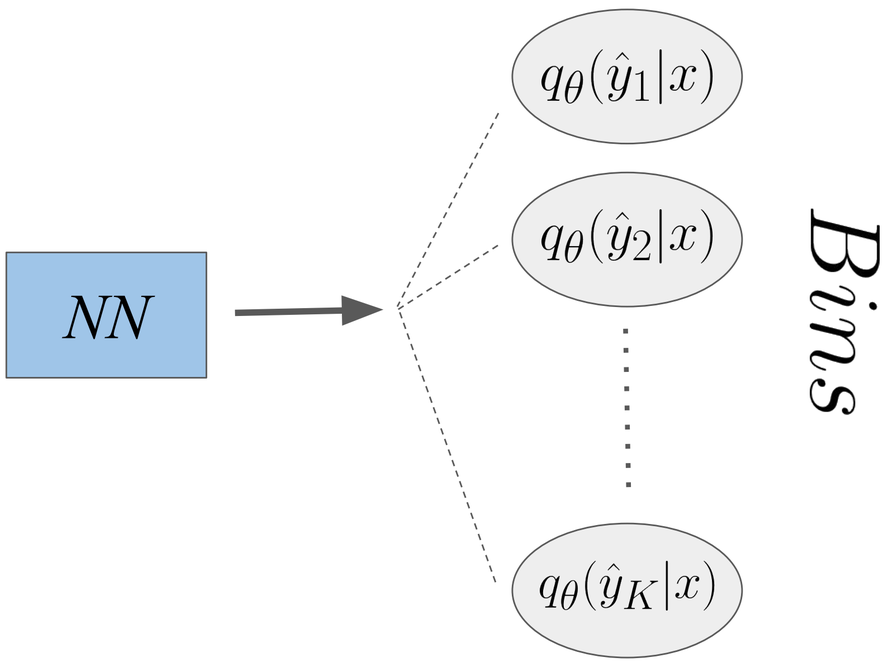
To find approximation $q_\theta \approx P$, we minimize the following expected loss function

$$\mathcal{L}(\theta) = \sum_{i=1}^{n_{\text{tr}}} \sum_{k=1}^{K} \ell(y_i, \hat{y}_k)\, q_\theta(\hat{y}_k \mid x_i) - \tau \mathcal{H}(q_\theta(\cdot \mid x_i)) \tag{1}$$

And convert regression into classification by binning the range space



(a) Binning the output space

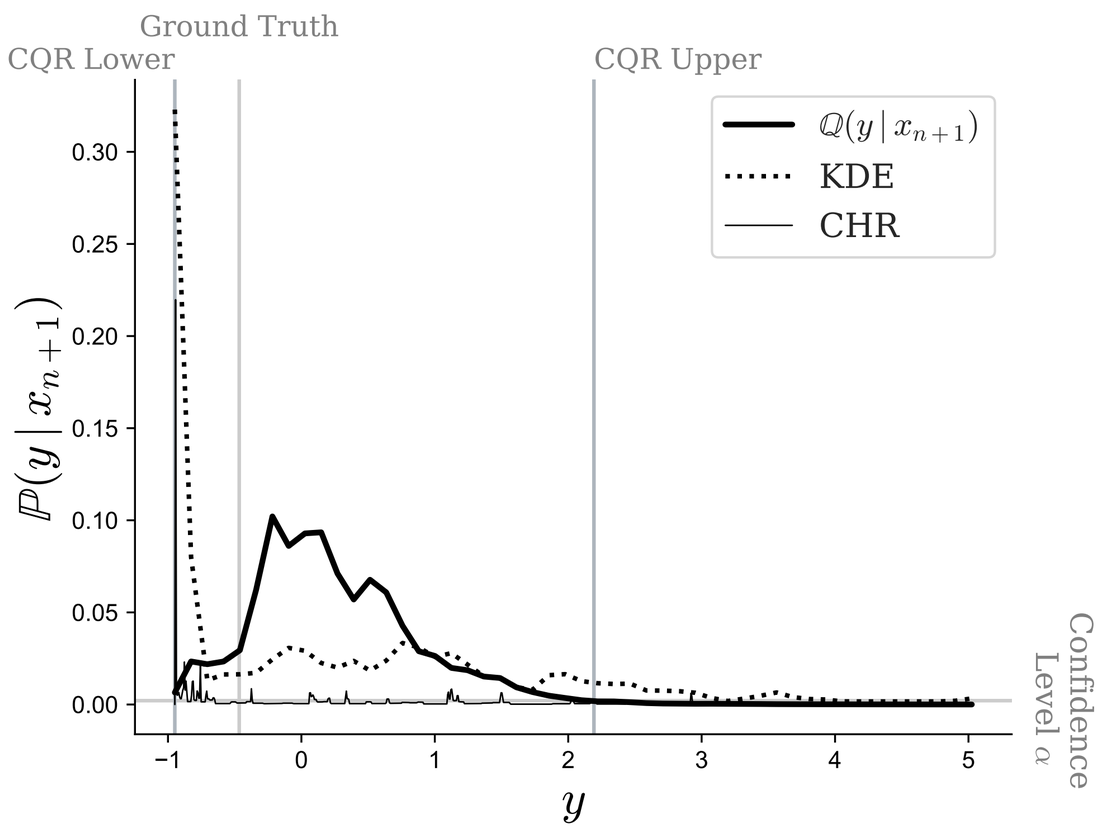


(b) Neural Network output space divided into K bins ($K = 50$ in our experiments)

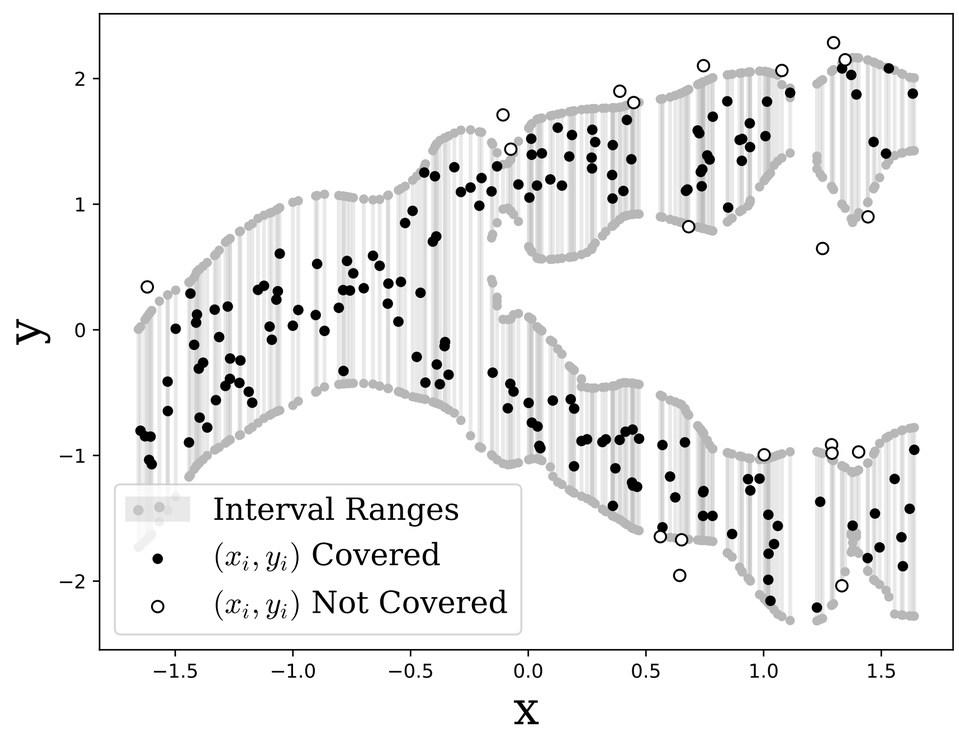
Experiments

Comparing model output lengths across datasets

Dataset	Stock	Cancer	Pendulum
CQR	$1.85_{(0.23)}$	$\mathbf{3.09_{(0.13)}}$	$2.25_{(0.31)}$
CB	$1.32_{(0.01)}$	$3.14_{(0.04)}$	$3.71_{(0.03)}$
CHR	$1.59_{(0.07)}$	$3.42_{(0.12)}$	$1.69_{(0.11)}$
R2CCP (Ours)	$\mathbf{0.92_{(0.02)}}$	$3.21_{(0.08)}$	$\mathbf{1.60_{(0.07)}}$



(a) Plot of outputted density functions for ours, KDE, and CHR on the Forest Dataset



(b) Our method can adaptively change the interval (shaded gray region) for the input values x across a variety of distributions