Distribution-free calibration guarantees for histogram binning without sample splitting

RECORDED: JUNE 20, 2021
PRESENTED: ICML WEEK, 2021

Carnegie Mellon University

Chirag Gupta (speaker)

Aaditya Ramdas

Calibration is a property of validity for probabilistic predictions

- Binary classification $Y \in \{0,1\}$, predictor $f: \mathcal{X} \to [0,1]$
- Calibration [mid 1900s meteorology literature, Dawid 1982]

$$P(Y = 1 \mid f(X)) = f(X)$$

- Meteorology: If the predicted chance of rain is $f(Day, Humidity, Wind, \dots) = 0.7$ on 10 days, then it rains on 7 days

- ML classification: If the predicted probability of class 1 is 0.7 on 10 test points, then 7 of them actually belong to class 1

Goal: $P(Y = 1 \mid f(X)) \approx f(X)$ without any distributional assumptions

Histogram binning is a popular *post-hoc* calibration method

Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers

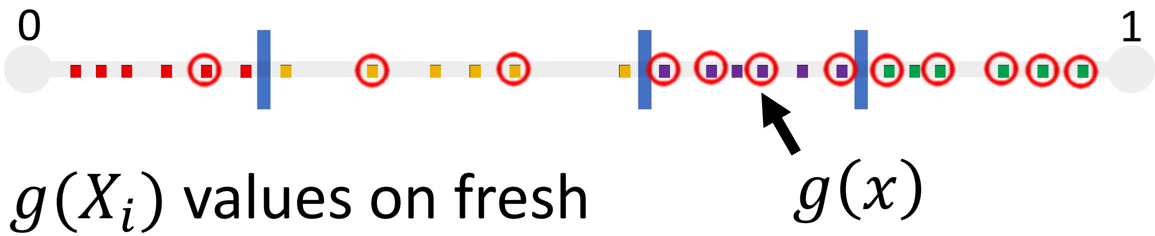
(ICML 2001)

Bianca Zadrozny ZADROZNY@CS.UCSD.EDU
Charles Elkan ELKAN@CS.UCSD.EDU

2.5 Calibrating naive Bayes classifier scores

We use a histogram method to obtain calibrated probability estimates from a naive Bayesian classifier. We sort the training examples according to their scores and divide the sorted set into b subsets of equal size, called bins. For each bin we compute lower and upper boundary $n(\cdot)$ scores. Given a test example x, we place it in a bin according to its score $n(x)$. We then estimate the corrected probability that x belongs to class j as the fraction of training examples in the bin that actually belong to j.

any pre-trained classifier $g\colon \mathcal{X} \to [0,1]$



$g(X_i)$ values on fresh calibration data

Predict $f(x) = 4/6$

| + ○ = Double dipping

We show tight calibration guarantees (such as high prob bounds on ECE) for histogram binning without sample splitting and without any distributional assumptions, using an elegant and powerful Markov property of order statistics

- Suppose Q is absolutely continuous wrt the Lebesgue measure and
  $Z_1, Z_2, Z_3 \sim Q$
  $Z_{(1)}, Z_{(2)}, Z_{(3)}$ are the order statistics

- Z_1, Z_2, Z_3 independent, but $Z_{(1)}, Z_{(2)}, Z_{(3)}$ dependent

- However,

$$Z_{(3)} \perp Z_{(1)} \mid Z_{(2)} \qquad \text{(!)}$$

- A first course in order statistics [SIAM, 2008]

What if I don’t care about theoretical guarantees?

- Based on the theory, we discuss how to choose the number of bins properly

- These principled choices lead to state-of-the-art performance on some deep learning problems: upcoming work being presented at ICML ‘21 workshops on Uncertainties in Deep Learning (Jul 23, pending acceptance) and Distribution-Free Uncertainty Quantification (Jul 24, spotlight talk)

- Easy-to-use code! https://github.com/aigen/df-posthoc-calibration

Distribution-free calibration guarantees for histogram binning without sample splitting

Carnegie Mellon University

Chirag Gupta and Aaditya Ramdas

https://github.com/aigen/df-posthoc-calibration

See you at the poster session and uncertainty-quantification workshops