DeepMind

Learning Optimal Conformal Classifiers


David Stutz

Krishnamurthy (Dj) Dvijotham

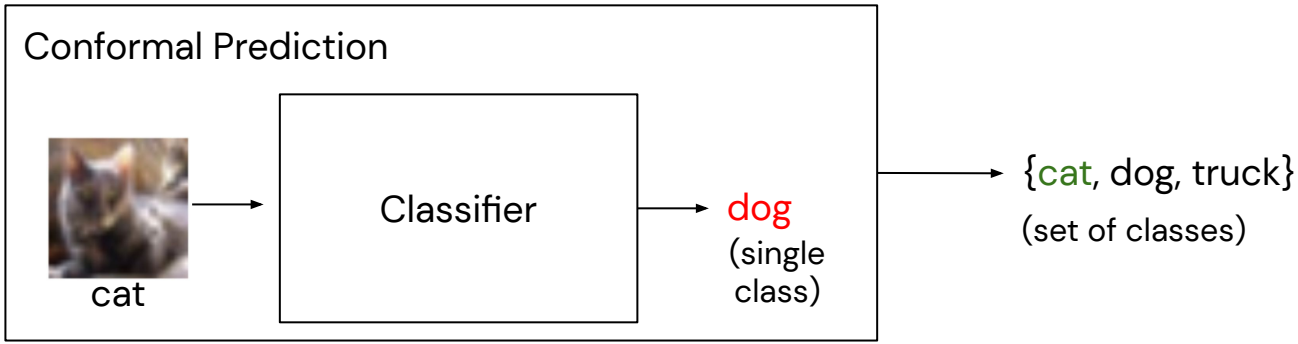

Ali Taylan Cemgil


Arnaud Doucet

ICLR 2022

Overview and Motivation: Conformal Training

Conformal prediction as post-training wrapper provides coverage guarantee:

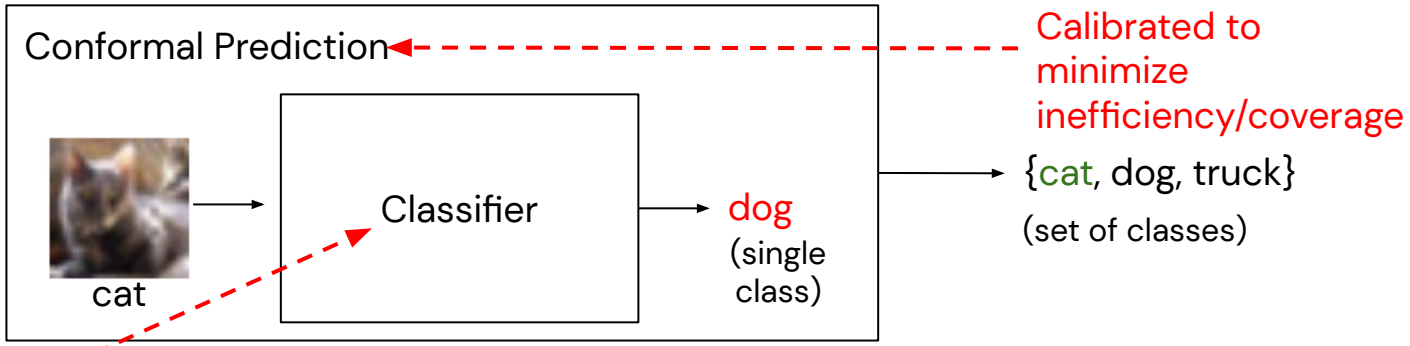


➔ True class is in the predicted confidence set with user-specified probability!

➔ Number of predicted classes = inefficiency

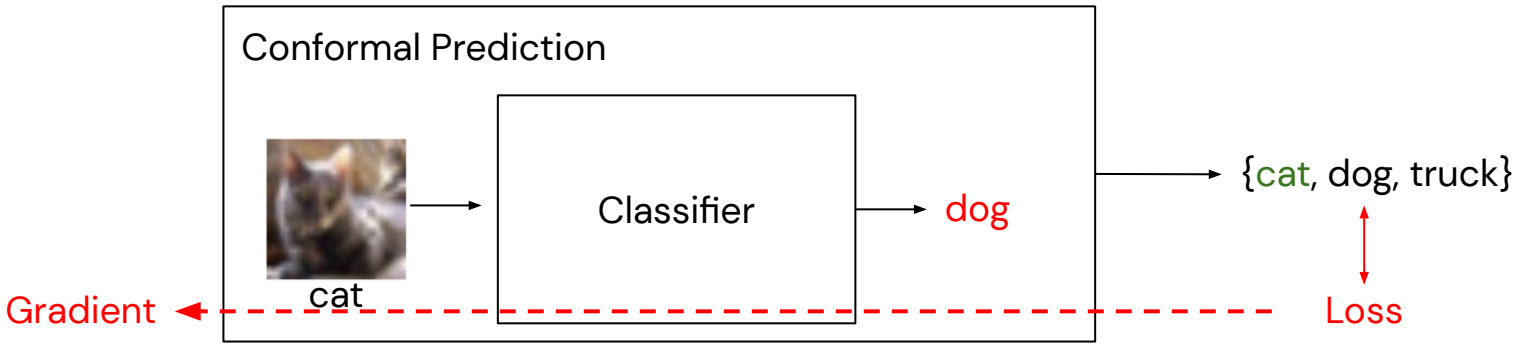
Overview and Motivation: Conformal Training

Conformal prediction as post-training wrapper provides marginal guarantee:



Overview and Motivation: Conformal Training

Conformal training = take conformal predictor into account during training:

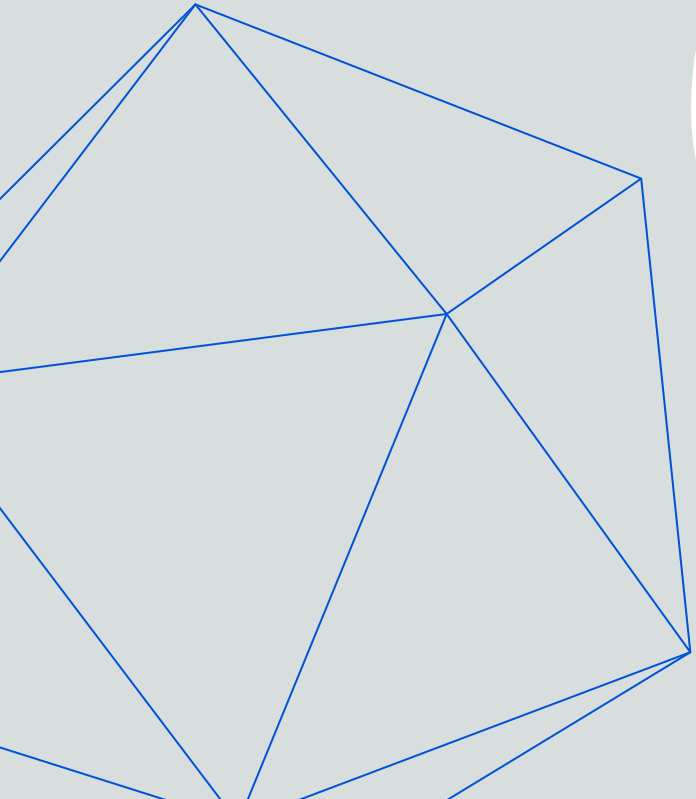


➔ Optimize arbitrary objectives defined on confidence sets

➔ Obtain guaranteed coverage using any conformal predictor after training.

DeepMind

Learning Optimal Conformal Classifiers



- ❏ Conformal Prediction
- ❏ Conformal Training
- ❏ Experimental Results
- ❏ Conclusion

Paper:
arxiv.org/abs/2110.09192

Conformal Prediction

For model $\pi_{\theta,y} \approx p(y|x)$, construct confidence sets $C_\theta(x) \subseteq [K] = \{1, \ldots, K\}$ such that:

$$P(y \in C_\theta(x)) \geq 1 - \alpha$$

- confidence level α user-specified

Conformal Prediction

For model $\pi_{\theta,y} \approx p(y|x)$, construct confidence sets $C_\theta(x) \subseteq [K] = \{1, \ldots, K\}$ such that:

$$P(y \in C_\theta(x)) \geq 1 - \alpha$$

- confidence level α user-specified
- *inefficiency* = average confidence set size $|C_\theta(x)|$ minimized

	{airplane}	{dog, cat}	{frog,horse,dog}	{cat,frog}	true class
coverage/inefficiency	yes/1	yes/2	no/3	yes/2	

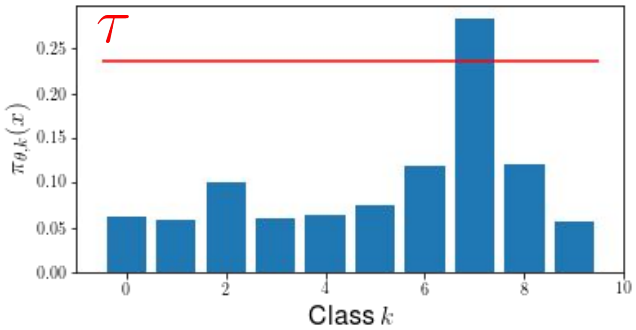
Example: Threshold Conformal Predictor

Two steps: *prediction* (test time) and *calibration* steps.

1. Prediction: define how confidence sets $C_\theta(x)$ are constructed,

$$C_\theta(x) := \{k \in [K] : E(x,k) := \pi_{\theta,k}(x) \geq \tau\}$$

with $E(x,k) := \pi_{\theta,k}(x)$ called conformity scores.



[1] Mauricio Sadinle, Jing Lei, and Larry Wasserman. Least ambiguous set-valued classifiers with bounded error levels. Journal of the American Statistical Association (JASA), 114(525):223–234, 2019.

Example: Threshold Conformal Predictor

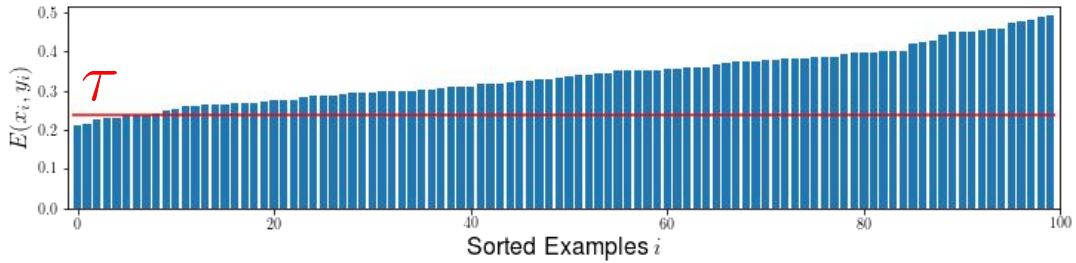
Two steps: *prediction* (test time) and *calibration* steps.

1. Prediction: define how confidence sets $C_\theta(x)$ are constructed.

$$C_\theta(x) := \{k \in [K] : E(x,k) := \pi_{\theta,k}(x) \geq \tau\}$$

2. Calibration: define threshold τ on held-out calibration set I_{cal}.

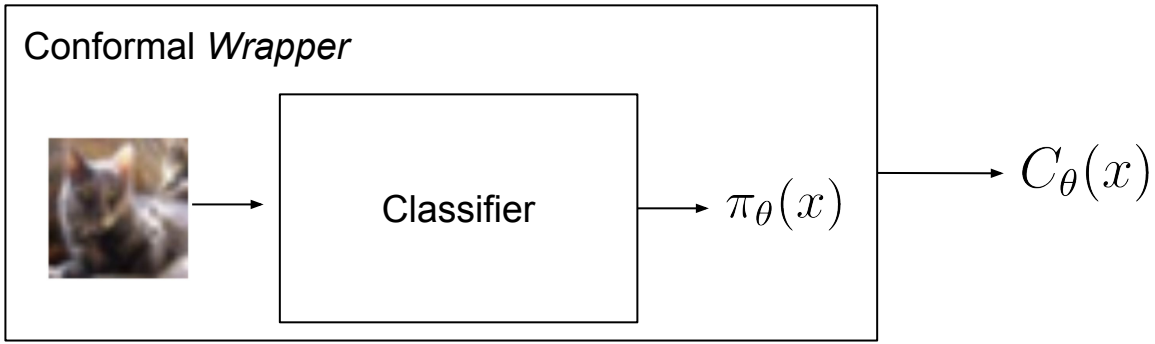
$\tau = \alpha$ -quantile of $\{E(x_i, y_i)\}_{i \in I_{\text{cal}}}$



[1] Mauricio Sadinle, Jing Lei, and Larry Wasserman. Least ambiguous set-valued classifiers with bounded error levels. Journal of the American Statistical Association (JASA), 114(525):223–234, 2019.

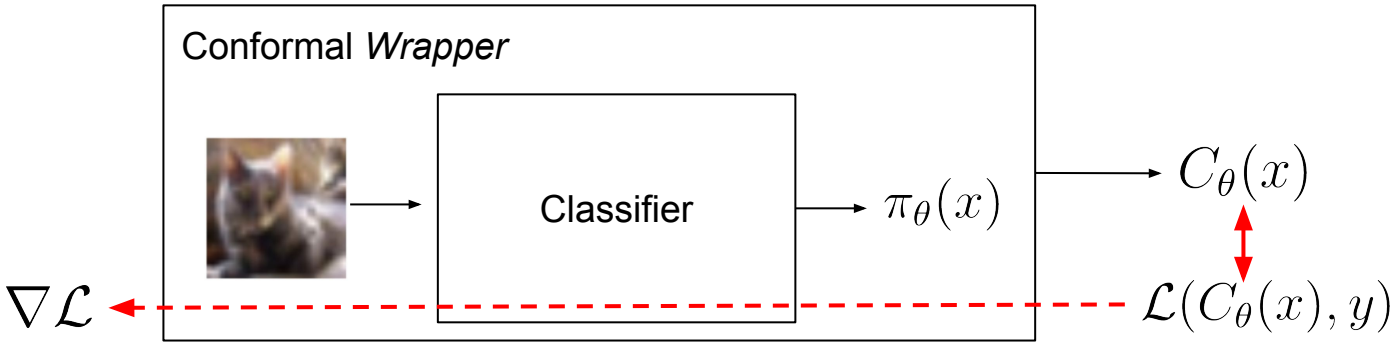
Training of Classifier *with* Conformal Wrapper

Conformal prediction is typically applied *after* training:



Training of Classifier *with* Conformal Wrapper

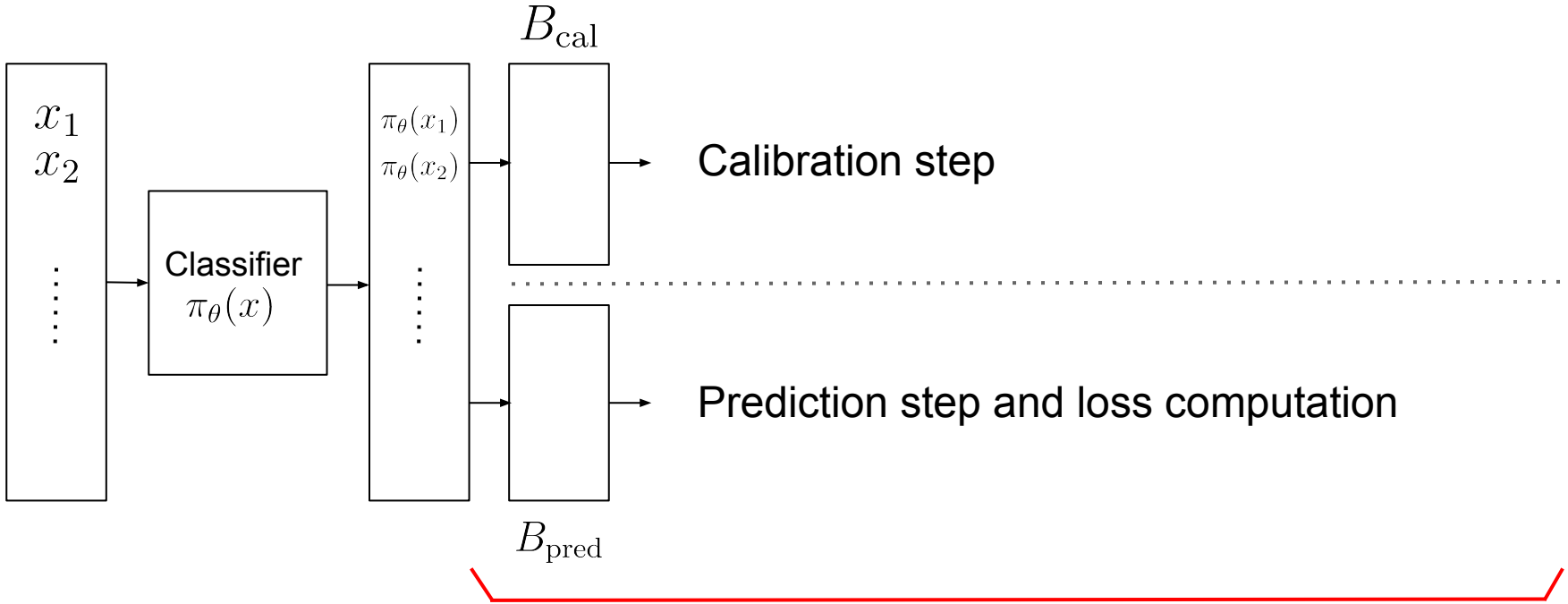
Conformal prediction is typically applied *after* training:



➔ Independent of conformal prediction method used at test time.

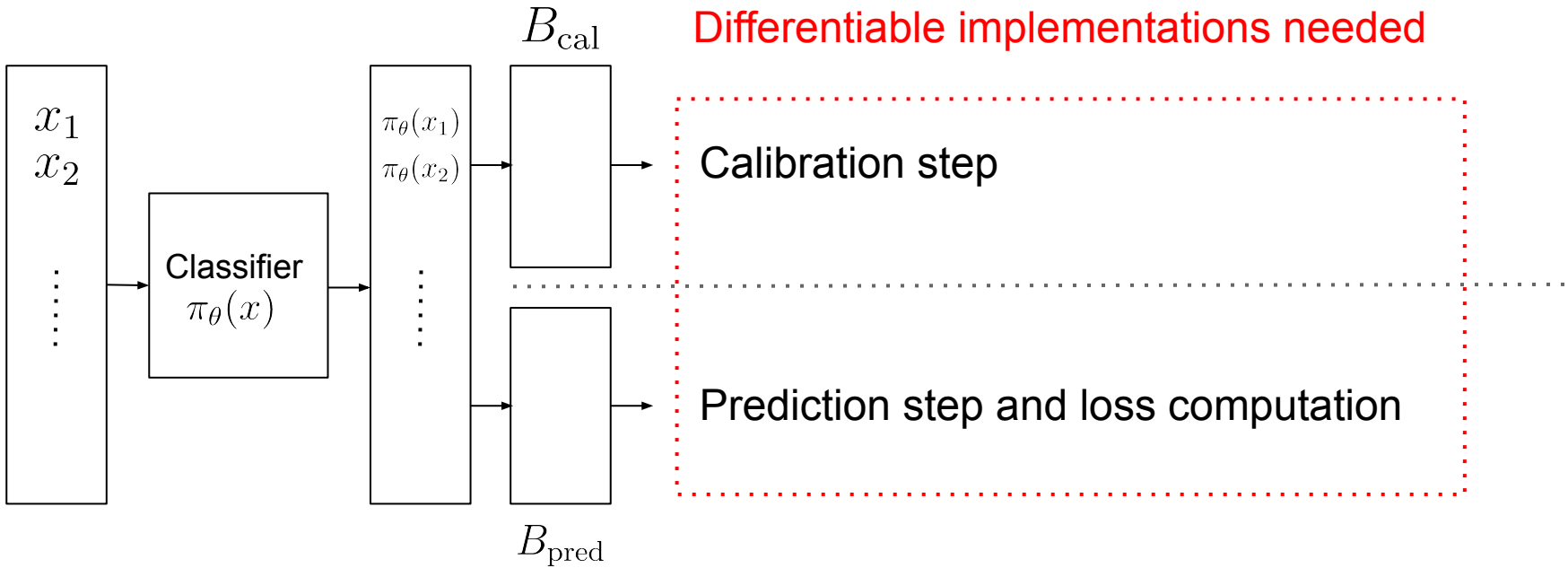
Conformal Training

Conformal Training

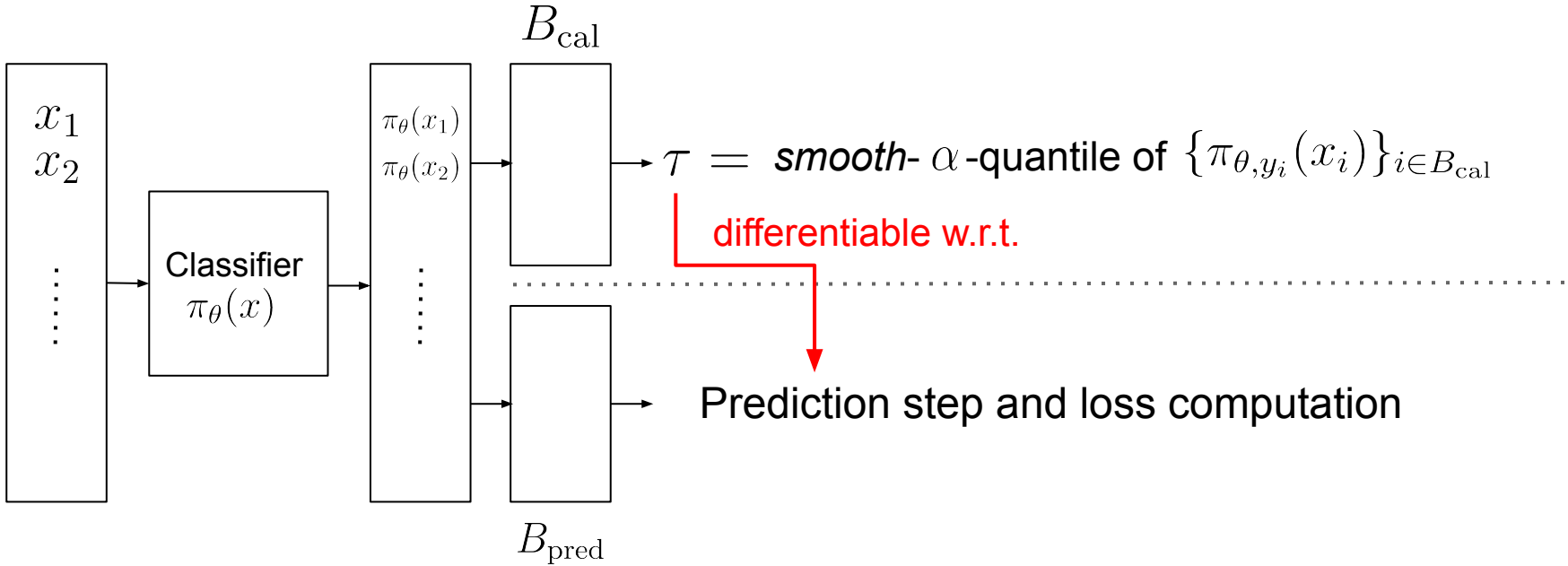


"Simulate" conformal prediction on each mini-batch

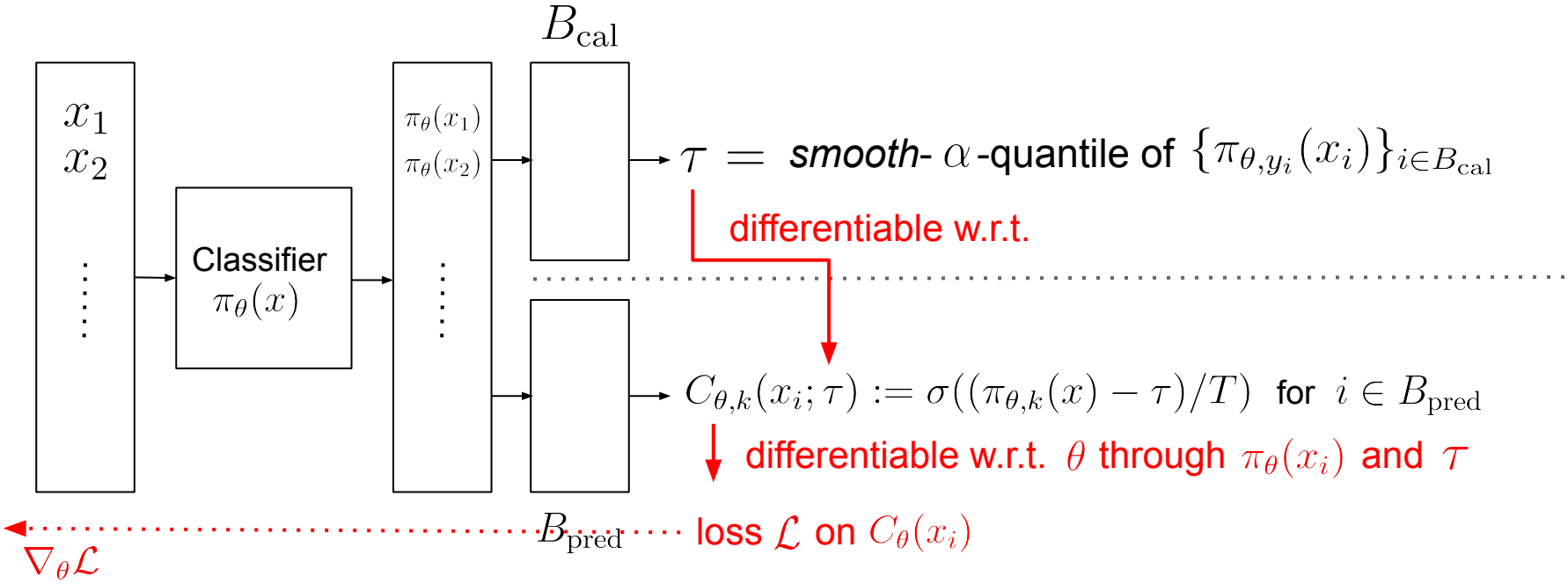
Conformal Training



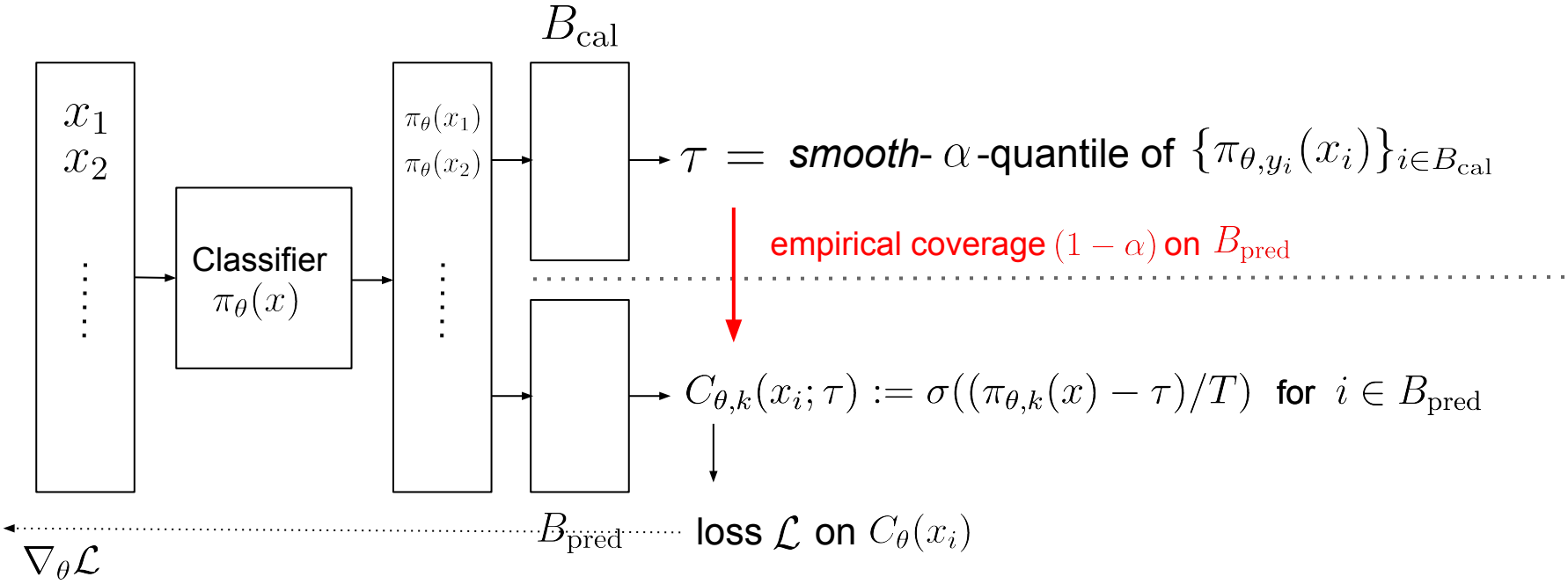
Conformal Training



Conformal Training



Conformal Training



➔ Re-calibrate at test time to obtain coverage guarantee!

Objectives

This talk:

- Reduce overall uncertainty
- Reduce *class-conditional* uncertainty

More applications in medical diagnosis in paper:

- Influence composition of confidence set

Optimizing Inefficiency

Train to directly reduce inefficiency:

$$\Omega(C_\theta(x)) = \sum_{k=1}^{K} C_{\theta,k}(x)$$

- $C_{\theta,k}(x) \in [0, 1]$ interpreted as “soft assignments”
- can be seen as smooth approximation of $\mathbb{E}[|C_\theta(x)|]$
- no loss on true label y as empirical coverage close to $(1 - \alpha)$

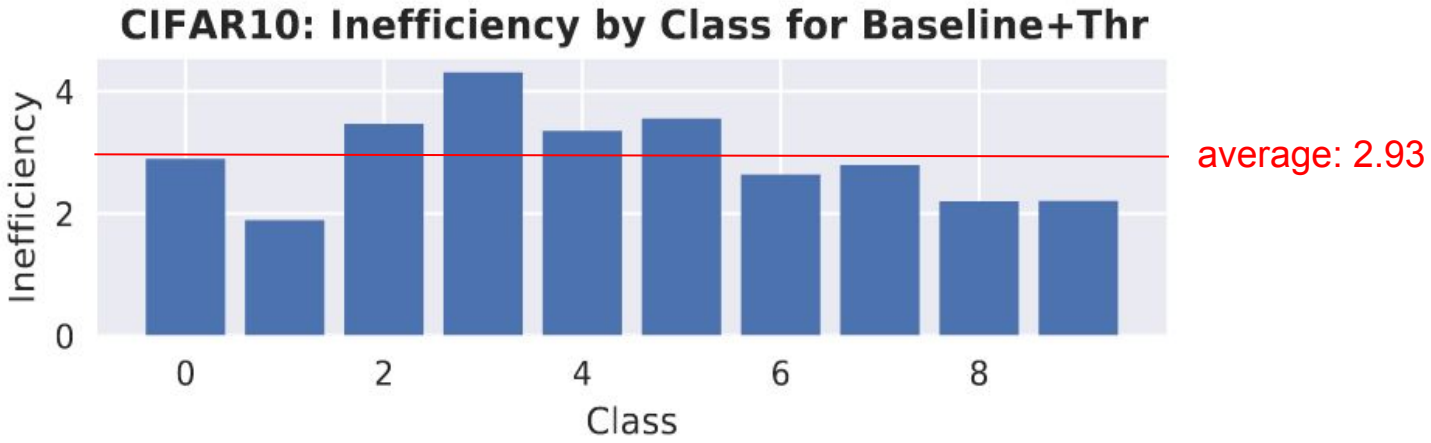
Reducing Inefficiency: Results

Inefficiency ↓ for $\alpha = 0.01$:				
CP at test time:	*Thr-Probs*		*APS [2]*	
Dataset	Baseline	Ours	Baseline	Ours
MNIST	2.23	**2.11** (-5.4%)	2.50	**2.14** (-14.14%)
F-MNIST	2.05	**1.67** (-18.5%)	2.36	**1.72** (-27.1%)
EMNIST (K = 52)	2.66	**2.49** (-6.4%)	4.23	**2.87** (-32.2%)
CIFAR10	2.93	**2.84** (-3.1%)	3.30	**2.93** (-11.1%)
CIFAR100	10.63	**10.44** (-1.8%)	16.62	**12.73** (-23.4%)

[2] Yaniv Romano, Matteo Sesia, and Emmanuel J. Candes. Classification with valid and adaptive coverage. In Advances in Neural Information Processing Systems (NIPS), 2020.

Inefficiency Distribution

Inefficiency ↓ distributed very differently across classes:



Results: CIFAR10

- Possible inefficiency improvement per class (in %)
- Cost in terms of average inefficiency increase across classes (in %)

Conclusion: Conformal Training

= end-to-end training of classifier and conformal wrapper.

- retains coverage guarantee
- reduces inefficiency
- allows arbitrary, application-specific losses

Paper: arxiv.org/abs/2110.09192