Active 🏃 Testing 🧪

OATML

UNIVERSITY OF OXFORD

OxCSML

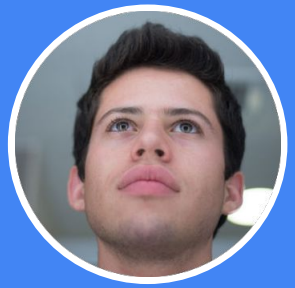
Sample–Efficent Model Evaluation

Jannik Kossen▽*

@janundnik

Sebastian Farquhar*

@seb_far



Yarin Gal

@yaringal

Tom Rainforth

@tom_rainforth

(▽ presenting, * equal

Active **Learning**

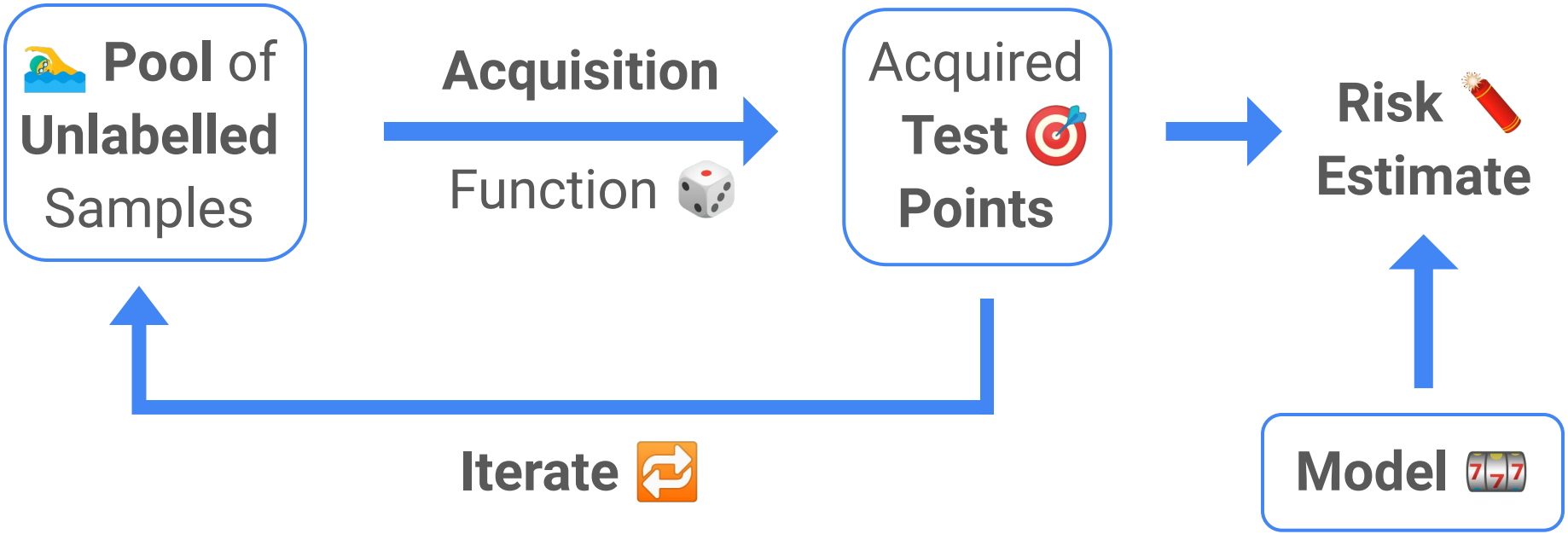
half of the solution

?

Active **Testing**

the **missing** half

Active Testing 📚 TL;DR 🤓



Acquisition Proposal

High loss points

contribute most to

model risk

$$q^*(i_m) \propto \mathbb{E}_{p(y|\boldsymbol{x}_{i_m})}\left[\mathcal{L}(f(\boldsymbol{x}_{i_m}), y)\right]$$

Active Testing – Kossen et al. – ICML 2021

👍 **Surrogate** Acquisition 👍

$$q(i_m) \propto \mathbb{E}_{\pi(\theta)\pi(y|\boldsymbol{x}_{i_m},\theta)}\left[\mathcal{L}(f(\boldsymbol{x}_{i_m}), y)\right]$$

Mean **Squared Error** Loss

$$q(i_m) \propto \left(f(\boldsymbol{x}_{i_m}) - \mathbb{E}_{\pi(y|\boldsymbol{x}_{i_m})}\left[y\right]\right)^2$$ **Disagreement**

$$+ \mathbb{V}_{\pi(\theta)}\left[\mathbb{E}_{\pi(y|\boldsymbol{x}_{i_m},\theta)}\left[y\right]\right]$$ **Epistemic Uncertainty**

$$+ \mathbb{E}_{\pi(\theta)}\left[\mathbb{V}_{\pi(y|\boldsymbol{x}_{i_m},\theta)}\left[y\right]\right]$$ **Aleatoric Uncertainty**

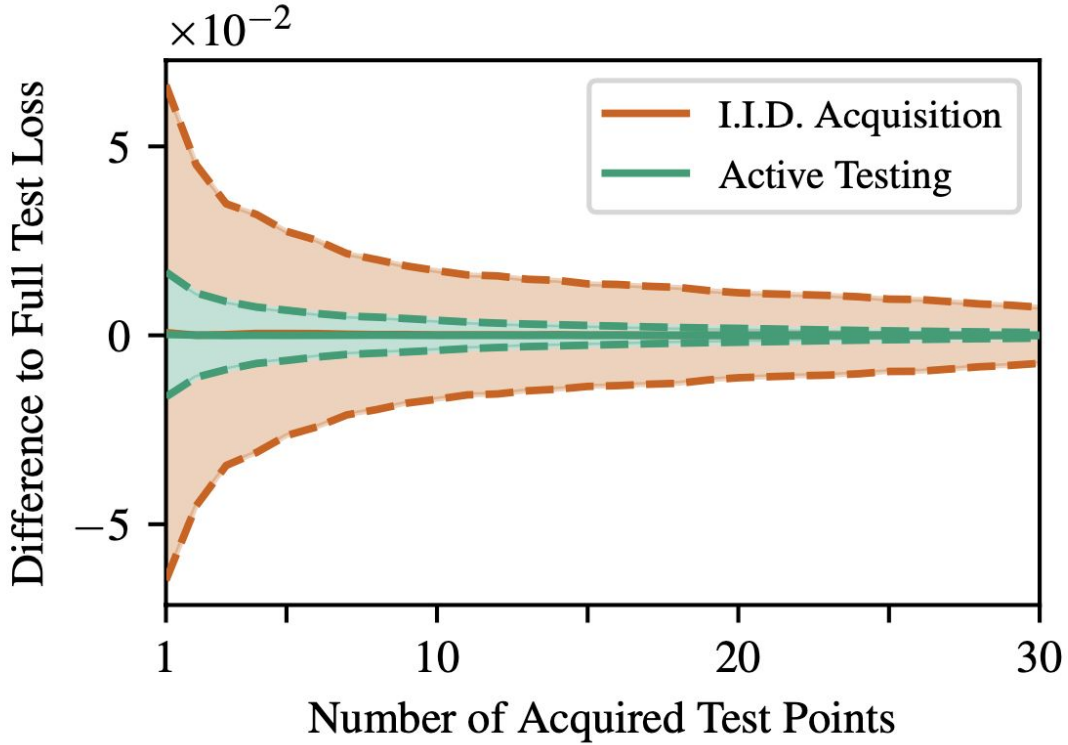
Why we need the **Surrogate**

1. **Extra Data**
2. **Diversity**

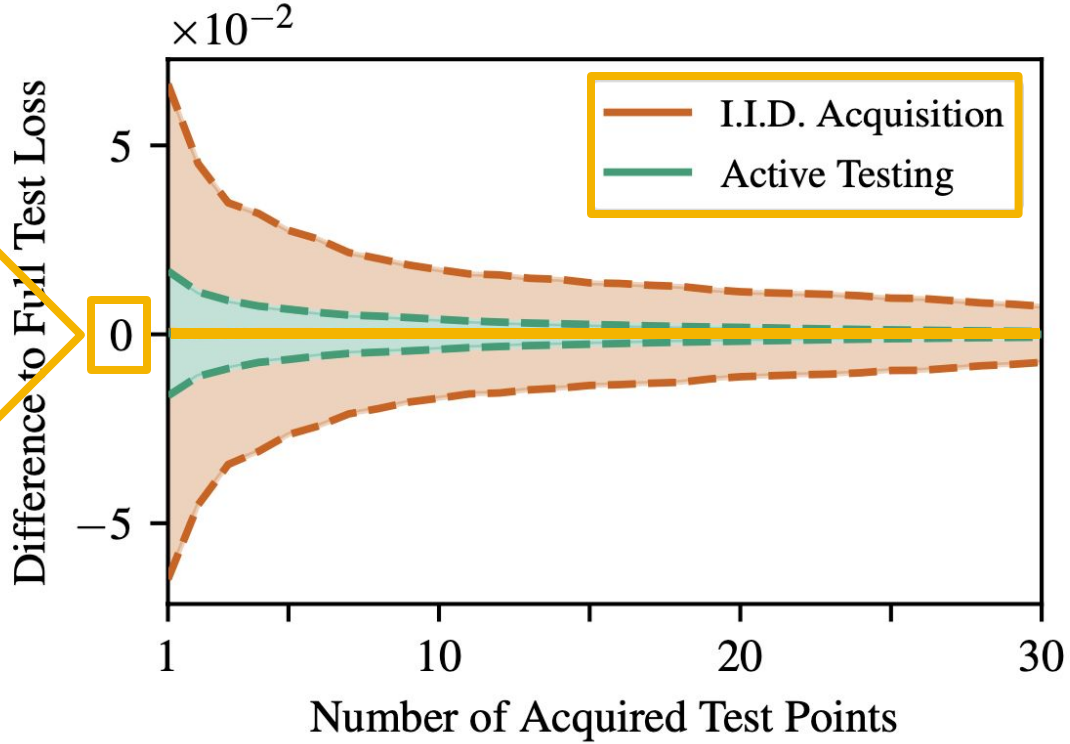
3 

Experiments

1


$\times 10^{-2}$
5
0
−5
Difference to Full Test Loss
I.I.D. Acquisition
Active Testing
1
10
20
30
Number of Acquired Test Points

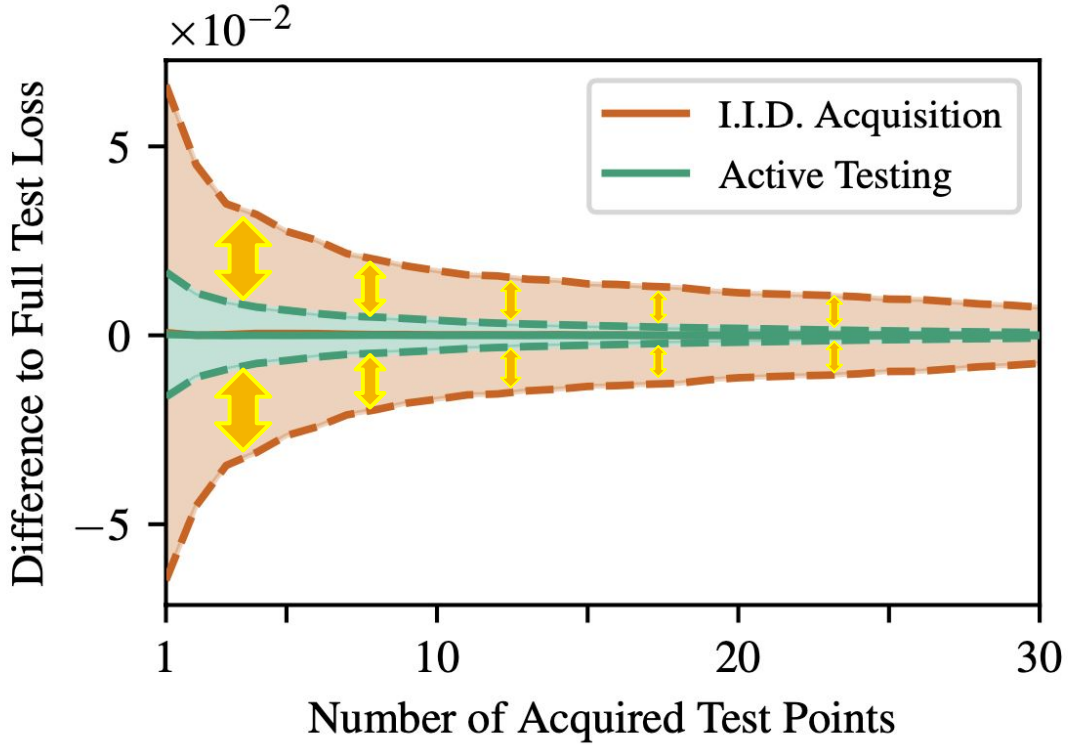
1



Active Testing
is
Unbiased

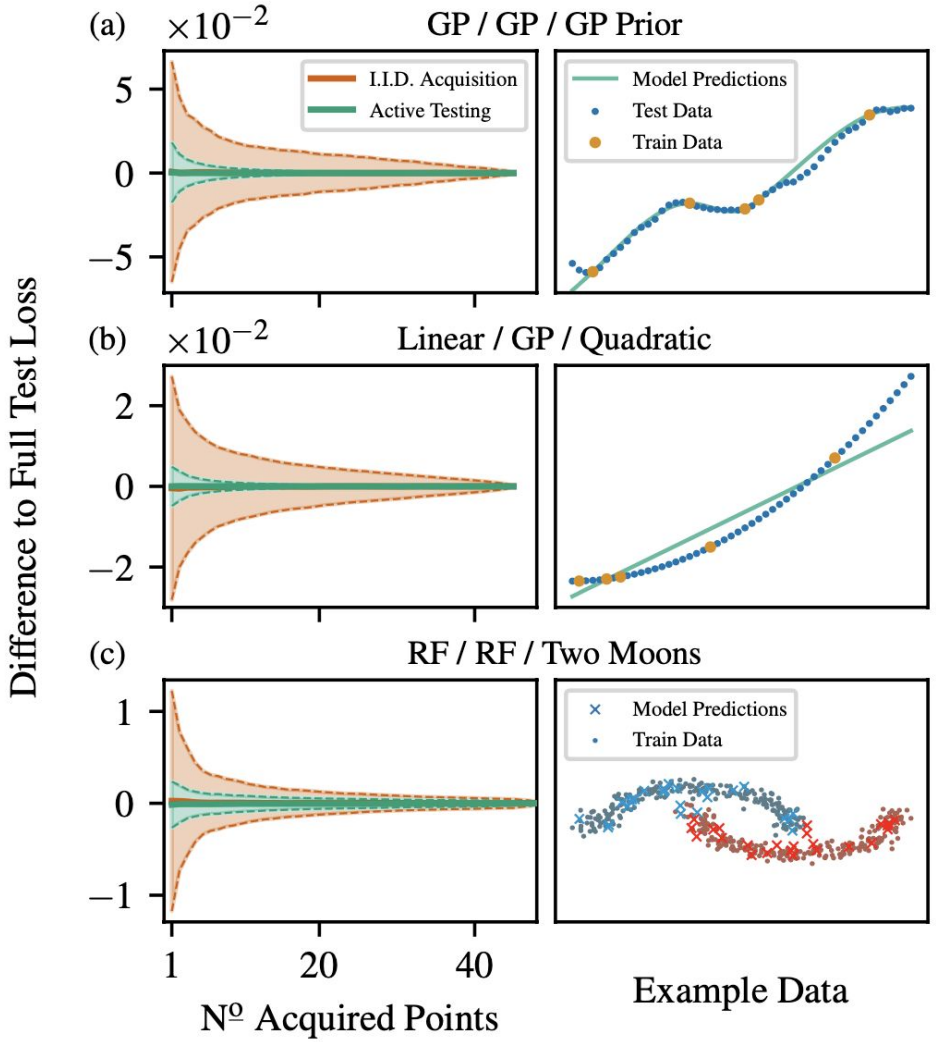
1

Active Testing
reduces
Variance

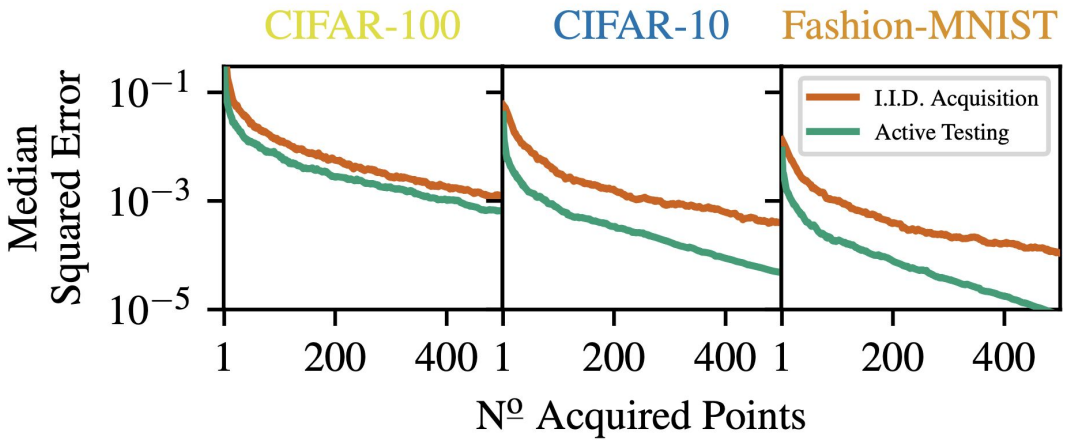


2

Gaussian Process
Random Forests
Linear Models
on
Toy Data



3



ResNets
on
Fashion-MNIST
CIFAR-10
CIFAR-100

That's **it**

Thanks **a lot**

Feel free to reach out with any ❓ you might have