ETH zürich

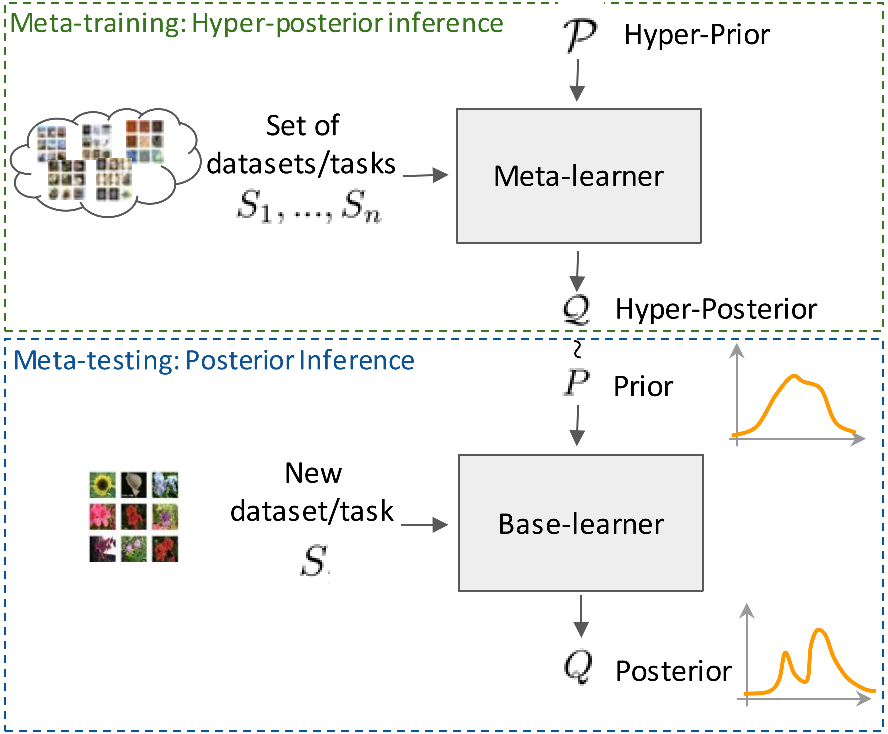
PACOH: Bayes-Optimal Meta-Learning with PAC-Guarantees

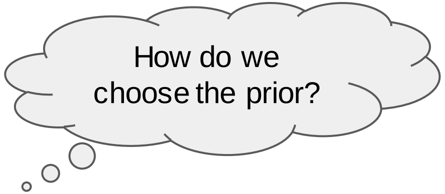



Jonas Rothfuss Vincent Fortuin Martin Josifoski Andreas Krause

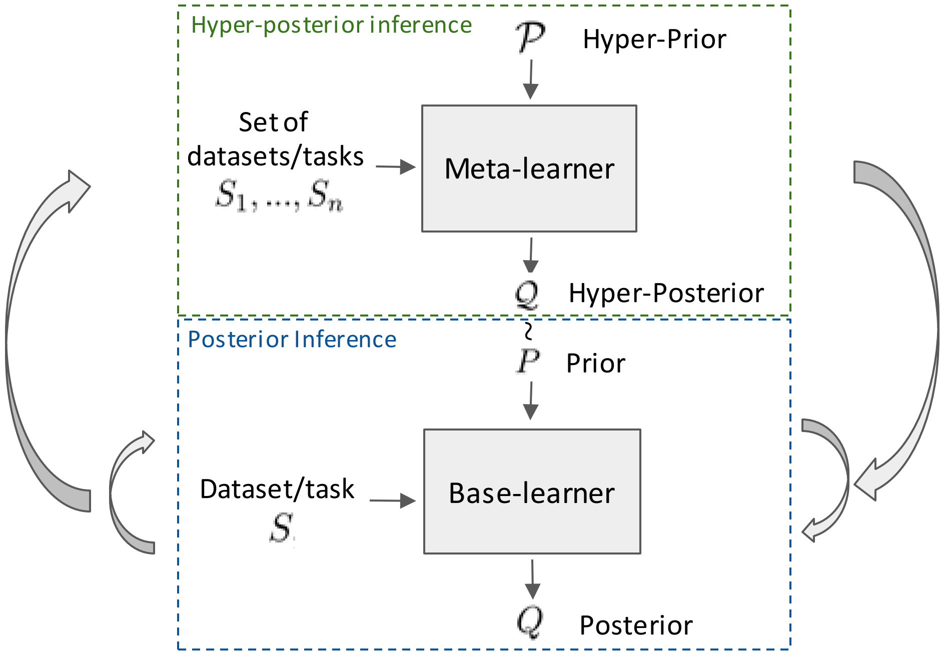
Link to the paper:
https://arxiv.org/abs/2002.05551

Meta-Learning PAC-Bayesian Priors





Meta-Learning Bayesian Priors



How do we know how good a prior is?

→ *For each task:*

1) *Run posterior inference*
2) *Evaluate posterior*

$$Q(S_i, P) = \arg\min_Q \hat{\mathcal{L}}(Q, S_i) + \frac{1}{\beta} KL(Q||P)$$

→ Optimization problem in itself!

PAC-Bayesian Meta-Learning
= Tricky two level optimization problem!
(e.g. Pentina and Lampert (2015), Amit and Meir (2018))

PAC-Bayesian Bound for Meta-Learning

Theorem 2. *(Informal) For all (hyper-posterior) distributions $\mathcal{Q}$, we have with probability at least $1-\delta$ that*

$$\underbrace{\mathcal{L}(\mathcal{Q},\mathcal{T})}_{\text{transfer error}} \le \frac{1}{n}\sum_{i=1}^{n} E_{P\sim\mathcal{Q}}\left[\underbrace{\hat{\mathcal{L}}(Q(S_i,P),S_i)}_{\text{empirical error}} + \frac{1}{\beta}\underbrace{KL\left(Q(S_i,P)||P\right)}_{\text{KL(posterior || prior)}}\right]$$

$$+ \underbrace{\left(\frac{1}{n\beta}+\frac{1}{\lambda}\right)KL(\mathcal{Q}||\mathcal{P})}_{\text{KL(hyper-posterior || hyper-prior)}} + C(\delta,\lambda,\beta)$$

What we care about

$$Q(S_i,P) = \arg\min_{Q} \hat{\mathcal{L}}(Q,S_i) + \frac{1}{\beta}KL(Q||P)$$

PAC-Bayesian Meta-Learning = Minimizing Upper Bound of Transfer Error

→ Tricky two level optimization problem!

(e.g. Pentina and Lampert (2015), Amit and Meir (2018))

The PAC-Optimal Hyper-Posterior (PACOH)

ETH zürich

Core contribution: **Closed form solution of the two-level optimization problem**!

Proposition 1. ***(PAC-Optimal Hyper-Posterior)*** *(informal) The hyper-posterior $\mathcal{Q}$ that minimizes the PAC-Bayesian meta-learning bound in Theorem 2 is given by*

$$\mathcal{Q}^*(P) = \frac{1}{Z^{II}} \mathcal{P}(P) \exp\left(\frac{1}{\sqrt{mn}+1} \sum_{i=1}^{n} \ln Z(S_i, P) \right)$$

wherein Z^{II} is a normalization constant.

(Generalized) Marginal Log-Likelihood

$$\ln Z(S_i, P_\phi) = \ln \mathbb{E}_{\theta \sim P_\phi} e^{-\beta \hat{\mathcal{L}}(\theta, S_i)}$$

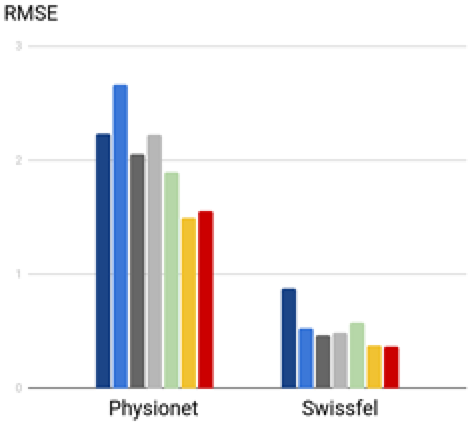
→ PAC-Bayesian meta-learning solution tractable up to normalization constant

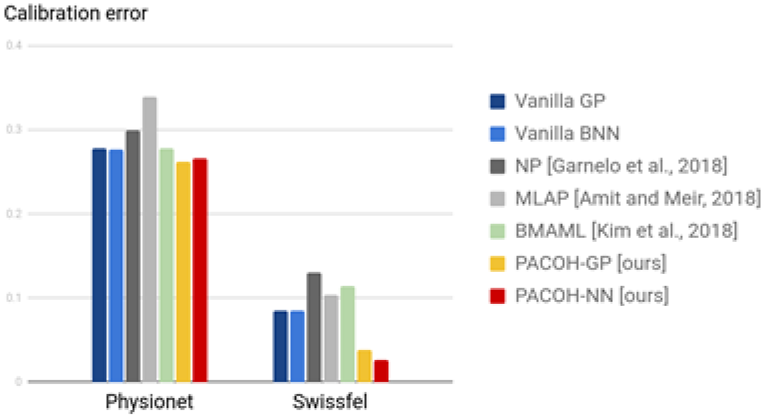
→ Now: Meta-Learning = Approximate inference on the PACOH

→ Scalable meta-learning of GP and NN priors

Benchmark study

Lower is better





Standard learners (no meta-learning)

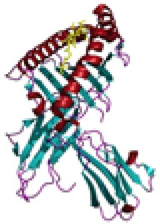
Previous meta-learners

PACOH (ours)

→ PACOH improves predictive accuracy

→ PACOH improves uncertainty estimates

Experiments - Meta-Learning BNN priors with PACOH

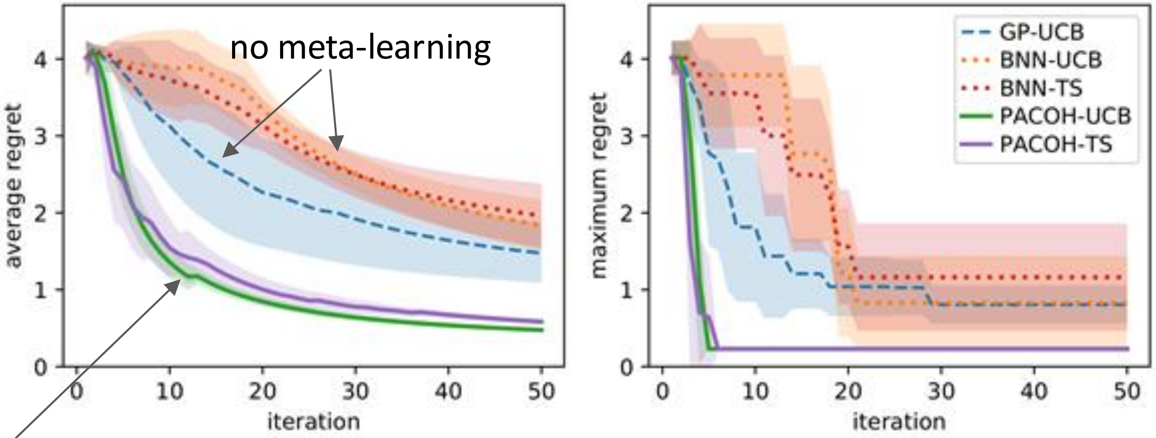


Meta-Learning for peptide-based vaccine development

1) **Bayesian Optimization:** Iteratively select molecule candidates to test for binding strength
2) **Meta-Learn BNN prior** with data of previous experiments for different surface protein alleles

Lower is better

no meta-learning



meta-learned PACOH-NN prior

Summary of Contributions

Core contribution - PAC-Bayesian Meta-Learning:

Previously: Two-level optimization problem

Now: Standard approximate inference on the PACOH

PACOH - A general meta-learning framework with

- Performance guarantees
- Principled meta-level regularization
- Uncertainty estimates

→ *Principled meta-learning algorithms for GPs and NNs*

PACOH: Bayes-Optimal Meta-Learning with PAC-Guarantees

Jonas Rothfuss, Vincent Fortuin, Andreas Krause

https://arxiv.org/abs/2002.05551