A Unified View on PAC-Bayes Bounds for Meta-Learning

Arezou Rezazadeh

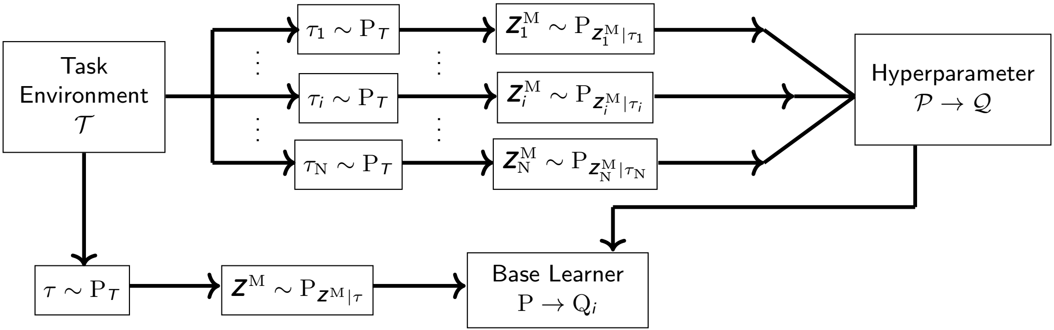
Chalmers University of Technology

2022 International Conference on Machine Learning

What Is Meta Learning?

- Process of automatically optimizing the hyperparameters
- Observe data from a number of related tasks
- Speed up the learning of a new, previously unseen task
- Active area of research
  - Meta-generalization gap
  - How to regularize the meta-learner, to avoid overfitting?
- This work:
  - A general framework that gives PAC-Bayes bounds
  - Re-obtaining classic, quadratic and fast-rate families
  - New PAC-Bayes classic bounds which reduce the meta-overfitting problem

Meta Learning



- Meta-generalization loss: $L(u) = \mathbb{E}_{P_{\mathcal{T}} P_{Z^M|\mathcal{T}}}\left[\mathbb{E}_{P_{W|Z^M,U=u}}[L_{P_{Z|\mathcal{T}}}(W)]\right]$
- Meta-training loss: $L_{Z_{1:N}^M}(u) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{P_{W|Z_i^M,U=u}}[L_{Z_i^M}(W)]$
- Meta-generalization gap: $\Delta L(u|Z_{1:N}^M) = L(u) - L_{Z_{1:N}^M}(u)$

Our Result

Theorem

Assume that $0 \leq \ell(\cdot,\cdot) \leq 1$; *then*

$$\mathrm{F}^{\mathsf{Env}}\left(\mathbb{E}_{U\sim\mathcal{Q}}\left(\mathrm{L}_{\mathrm{P}_{T\boldsymbol{z}^{\mathrm{M}}}}(U)\right), \mathbb{E}_{U\sim\mathcal{Q}}\left(\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\tilde{\mathrm{L}}_{\boldsymbol{z}_i^{\mathrm{M}}}^{T_i}(U)\right)\right)$$
$$+\,\mathrm{F}^{\mathsf{Task}}\left(\mathbb{E}_{U\sim\mathcal{Q}}\left(\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\tilde{\mathrm{L}}_{\boldsymbol{z}_i^{\mathrm{M}}}^{T_i}(U)\right), \mathbb{E}_{U\sim\mathcal{Q}}\left(\mathrm{L}_{\boldsymbol{z}_{1:\mathrm{N}}^{\mathrm{M}}}(U)\right)\right)$$
$$\leq \left(\frac{1}{\theta_{\mathsf{tsk}}}+\frac{1}{\theta_{\mathsf{env}}}\right)\mathrm{D}_{\mathsf{KL}}\left(\mathcal{Q}||\mathcal{P}\right)+\frac{1}{\mathrm{N}\cdot\theta_{\mathsf{tsk}}}\mathbb{E}_{\mathcal{Q}}\left(\sum_{i=1}^{\mathrm{N}}\mathrm{D}_{\mathsf{KL}}\left(\mathrm{Q}_i||\mathrm{P}\right)\right)+c$$

- Re-obtaining existing bounds: $\sqrt{\mathrm{D}_{\mathsf{KL}}(\mathcal{Q}||\mathcal{P})}+\sqrt{\mathrm{D}_{\mathsf{KL}}(\mathrm{Q}_i||\mathrm{P})}$
- New classic bound: merge two terms: $\sqrt{\mathrm{D}_{\mathsf{KL}}(\mathcal{Q}||\mathcal{P})+\mathrm{D}_{\mathsf{KL}}(\mathrm{Q}_i||\mathrm{P})}$