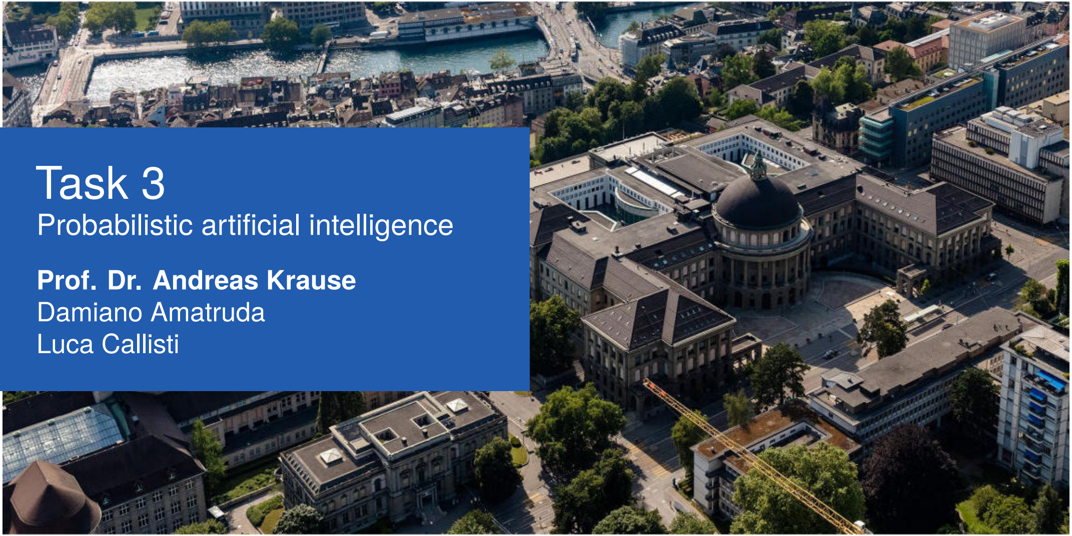
ETH zürich



Task 3

Probabilistic artificial intelligence

Prof. Dr. Andreas Krause
Damiano Amatruda
Luca Callisti

Constrained Bayesian Optimization

Constrained Bayesian Optimization consists in solving the problem:

$$x^* \in \underset{x \in \mathcal{X}, v(x) < \kappa}{\arg\max} \ f(x),$$

by observing noisy values of an objective function f and a constraint function v.

We approximate these with Gaussian Process models and we sample by iteratively optimizing an *acquisition function*, which in our case we define as:

$$\mathrm{AF}(x) = \begin{cases} \Pr(v_{\mathrm{GP}}(x) < \kappa)\, \mathrm{EI}(x) & \text{if } \Pr(v_{\mathrm{GP}}(x) < \kappa) \geq 1 - \delta, \\ 0 & \text{otherwise,} \end{cases}$$

where δ is the tolerance.

We use the *expected improvement*, defined as:

$$\mathrm{EI}(x) = \mathbb{E}_{f_{\mathrm{GP}}(x)}[\max(0, f_{\mathrm{GP}}(x) - f^*)],$$

where f^* is the running optimum, since it has a good balance between exploration and exploitation.

Plug-In Estimate for Running Optimum

To adapt the expected improvement to the noisy environment, we use a plug-in estimate for the running optimum:

$$\mathrm{EI}(x) = \mathbb{E}_{f_{\mathrm{GP}}(x)}\left[\max\left(0, f_{\mathrm{GP}}(x) - \max_{x\in\mathcal{D},\mu_{v_{\mathrm{GP}}}(x)<\kappa} \mu_{f_{\mathrm{GP}}}(x) - \xi\right)\right],$$

where:

- $\mathcal{D}$ is the set of previous sampled points;
- ξ determines how much exploration is done, pretending to have a higher running optimum than we really have.

LogEI

To avoid numerical instabilities in the computation of EI, we optimize the logarithm of the acquisition function using LogEI, which can be expressed in closed form as:

$$\mathrm{LogEI}(x) = \log \sigma_{f_{\mathrm{GP}}}(x) + \log h(z),$$

where $z = \left(\mu_{f_{\mathrm{GP}}}(x) - \max_{x \in \mathcal{D}} \mu_{f_{\mathrm{GP}}}(x) - \xi\right) / \sigma_{f_{\mathrm{GP}}}(x)$ and $h(z) = \phi(z) + z\Phi(z)$.

We compute $\log h(z)$ as:

$$\log h(z) = \begin{cases} -\frac{z^2}{2} - \log(2\pi) - 2\log|z| & \text{if } z \leq -1/\sqrt{\epsilon}, \\ -\frac{z^2}{2} - \log(2\pi) + \operatorname{log1mexp}\left(\log\left(\operatorname{erfcx}\left(-\frac{z}{\sqrt{2}}\right)|z|\right) + \frac{\log(\frac{\pi}{2})}{2}\right) & \text{if } -1/\sqrt{\epsilon} < z \leq -1, \\ \log(\phi(z) + z\Phi(z)) & \text{if } z > -1, \end{cases}$$

$$\operatorname{log1mexp}(z) = \log(1 - \exp(z)) = \begin{cases} \log(-\operatorname{expm1}(z)) & \text{if } z \geq -\log 2, \\ \operatorname{log1p}(-\exp(z)) & \text{if } z < -\log 2. \end{cases}$$

where ϵ is the floating point resolution.

Bibliography

[1] Michael A. Gelbart, Jasper Snoek, and Ryan P. Adams. *Bayesian Optimization with Unknown Constraints*. 2014. arXiv: 1403.5607 [stat.ML]. URL: https://arxiv.org/abs/1403.5607.

[2] Ziyu Wang and Nando de Freitas. *Theoretical Analysis of Bayesian Optimisation with Unknown Gaussian Process Hyper-Parameters*. 2014. arXiv: 1406.7758 [stat.ML]. URL: https://arxiv.org/abs/1406.7758.

[3] Sebastian Ament et al. *Unexpected Improvements to Expected Improvement for Bayesian Optimization*. 2024. arXiv: 2310.20708 [cs.LG]. URL: https://arxiv.org/abs/2310.20708.

[4] Martin Mächler. *Accurately Computing log(1 - exp(- |a|)) Assessed by the Rmpfr package*. Sept. 2015. DOI: 10.13140/RG.2.2.11834.70084.