Bayesian Active Learning by Soft Mean Objective Cost of Uncertainty

Guang Zhao[1], Edward R. Dougherty[1], Byung-Jun Yoon[1,3], Francis J. Alexander[3], Xiaoning Qian[1,2]

[1]Electrical & Computer Engineering, [2]Computer Science & Engineering, Texas A&M University, [3]Brookhaven National Laboratory

AISTATS 2021
Poster represented on Apr. 13 at 14:00-16:00

Myopic acquisition function may get stuck

- Active learning sequentially selects training samples for labelling by optimizing an acquisition function.

- We focus on acquisition function based on Mean Objective Cost of Uncertainty (MOCU) :
  - MOCU measures the direct influence on classification error due to the model uncertainty.
  - As an acquisition function, *MOCU reduction* is optimal for one-step queries.
  - However, it may get stuck before converging to the optimal classifier.

MOCU

- Consider model $p(y|x,\theta)$ with uncertain parameters $\theta \sim \pi(\theta)$
- MOCU is a piecewise linear function of $\pi(\theta)$:

$$\mathcal{M}\big(\pi(\theta)\big) = \mathbb{E}_{\boldsymbol{x}}\Big\{\underbrace{\mathbb{E}_{\pi(\theta)}\big[\max_y p(y|\boldsymbol{x},\theta)\big]}_{\text{Linear function}} - \underbrace{\max_y p(y|\boldsymbol{x})}_{\text{Piecewise linear function}}\Big\}.$$

- *MOCU reduction*:

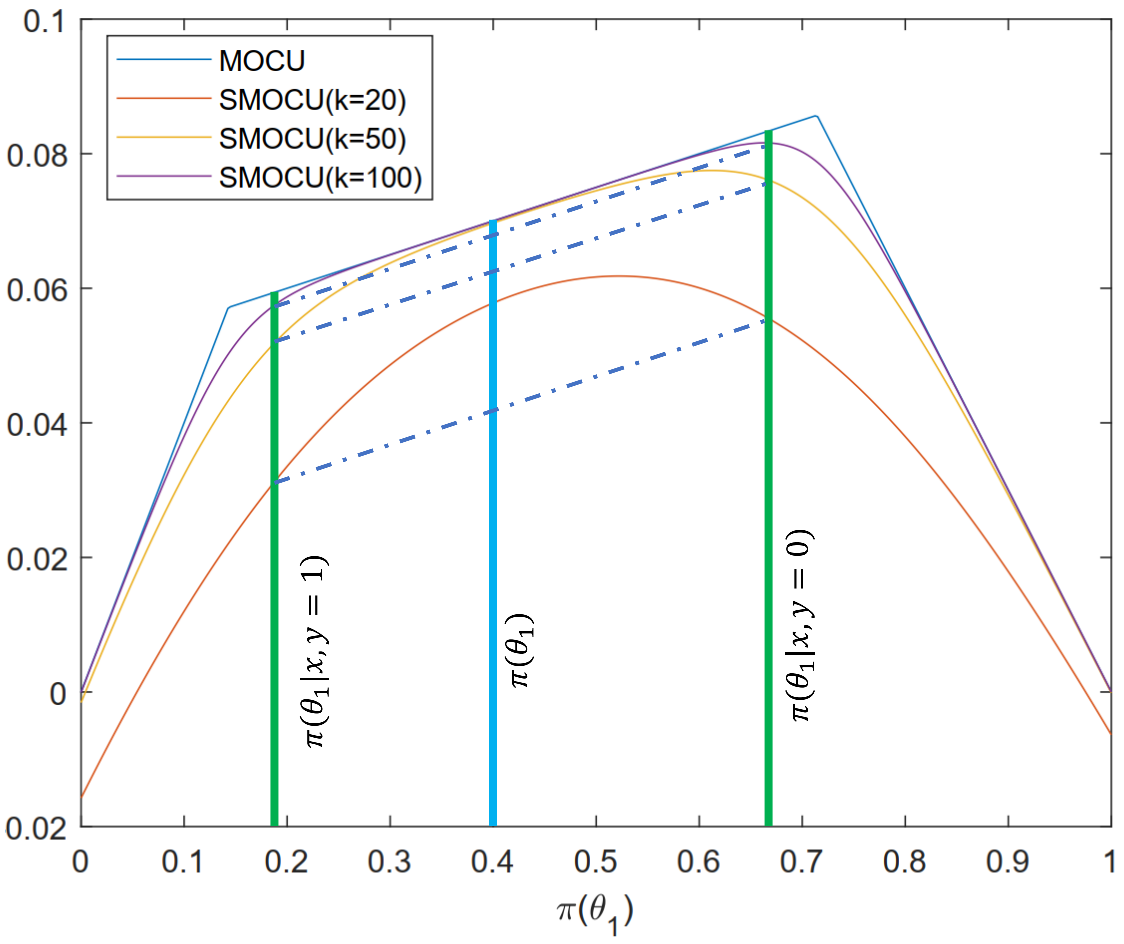
$$U^M\big(\boldsymbol{x};\pi(\theta)\big) = \mathcal{M}\big(\pi(\theta)\big) - \mathbb{E}_{p(y|\boldsymbol{x})}\big[\mathcal{M}\big(\pi(\theta|\boldsymbol{x},y)\big)\big].$$

MOCU Analysis

- Due to the piecewise linearity of MOCU, the *MOCU reduction* can be 0.

$$U^M(\boldsymbol{x};\pi(\theta)) = \mathcal{M}(\pi(\theta)) - \mathbb{E}_{p(y|\boldsymbol{x})}[\mathcal{M}(\pi(\theta|\boldsymbol{x},y))]$$

- As a result, *MOCU reduction* may not capture the long-term effect of querying $\boldsymbol{x}$.



Soft MOCU

- SMOCU approximates MOCU with a smooth strictly concave function.

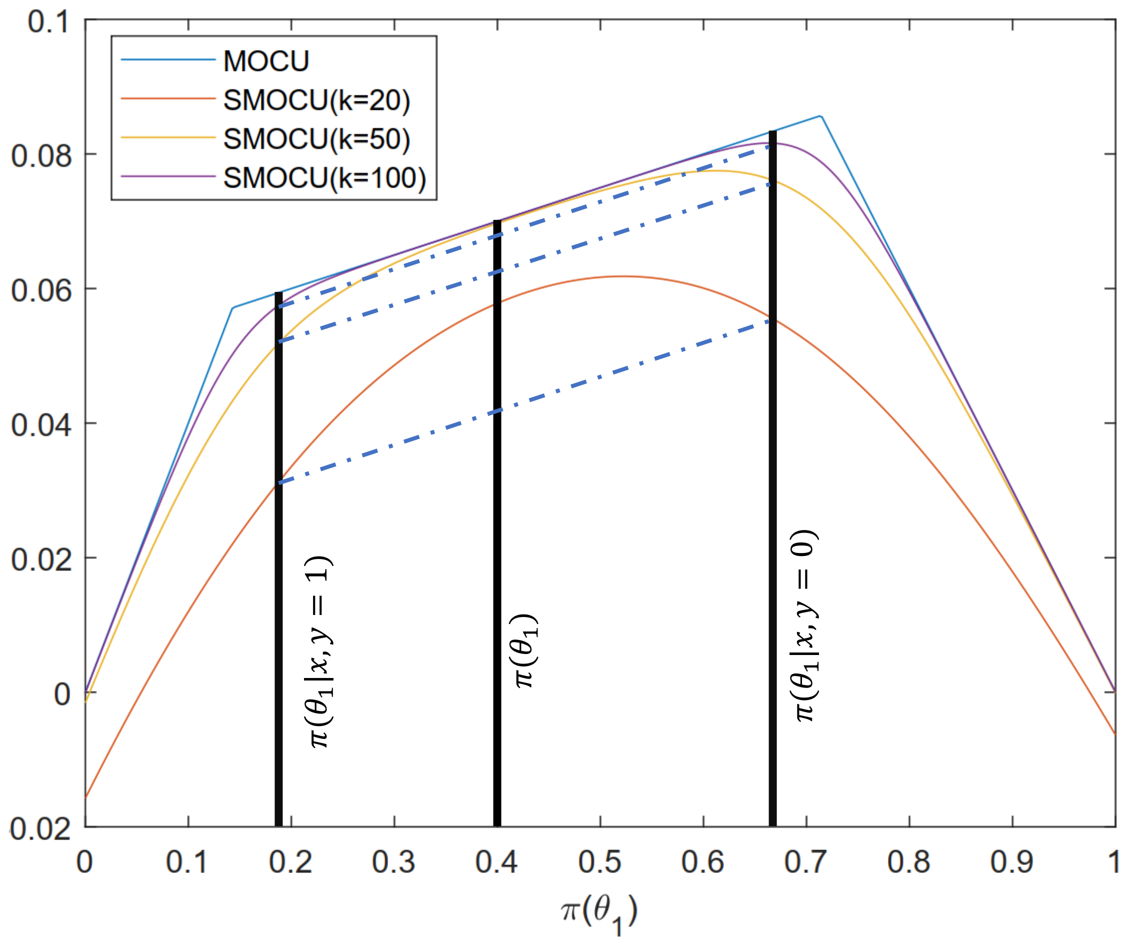
$$\mathcal{M}(\pi(\theta)) = \mathbb{E}_{\boldsymbol{x}}\left\{\mathbb{E}_{\pi(\theta)}[\max_y p(y|\boldsymbol{x},\theta)] - \max_y p(y|\boldsymbol{x})\right\}$$

$$\mathcal{M}^s(\pi(\theta)) = \mathbb{E}_{\boldsymbol{x}}\left\{\mathbb{E}_{\pi(\theta)}[\max_y p(y|\boldsymbol{x},\theta)] - \frac{1}{k}\log\left[\sum_y \exp\left(k \cdot p(y|\boldsymbol{x})\right)\right]\right\}$$

- *SMOCU reduction*:

$$U^s(\boldsymbol{x};\pi(\theta)) = \mathcal{M}^s(\pi(\theta)) - \mathbb{E}_{p(y|\boldsymbol{x})}[\mathcal{M}^s(\pi(\theta|\boldsymbol{x},y))].$$

  - Near optimal performance in short-term
  - Avoid myopic behavior in the long run



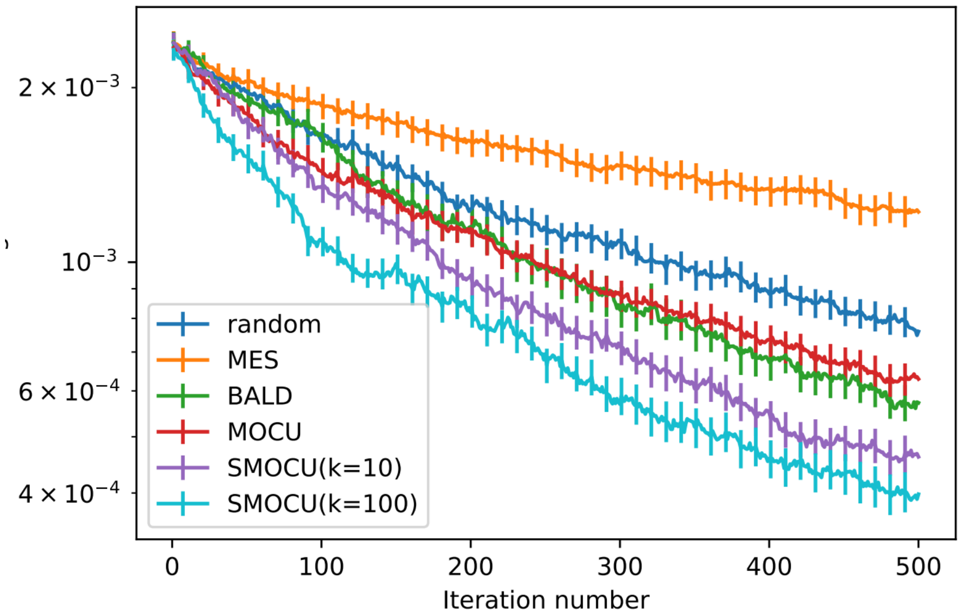
Theoretical Result

Theorem: Assume that in a classification problem, both $\mathcal{X}$ and Θ are discrete with finite elements, the true model parameter $\theta_r \in \Theta$ and the prior $\pi^0(\theta_r) > 0$; then for the active learning algorithm defined by the acquisition function $U^s(\boldsymbol{x}; \pi^n(\theta))$, we have

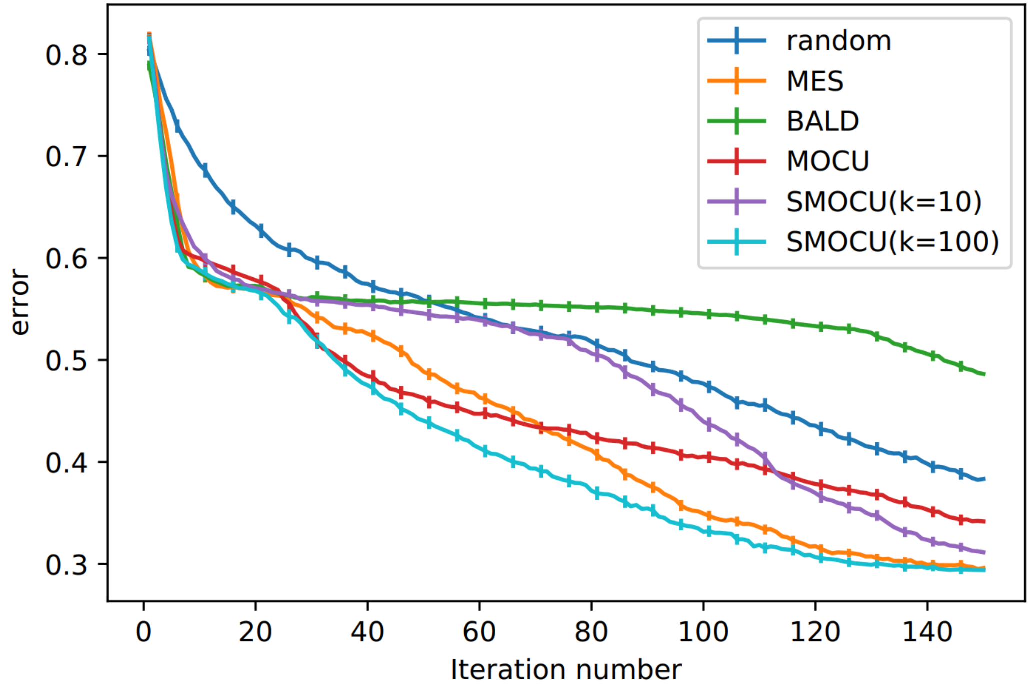
$$\mathcal{M}\left(\pi^n(\theta)\right) \xrightarrow{\text{a.s.}} 0 \text{ as } n \to \infty,$$

Indicating that the active learning algorithm approaches the optimal classifier almost surely.

Experiment results



2d Bayesian logistic regression model with an error box in the center



Classification error comparison on *UCI User Knowledge* dataset.

Code available at https://github.com/QianLab/Soft_MOCU

Summary of the study

- We propose a novel acquisition function, *Soft MOCU (SMOCU) reduction*, for pool-based Bayesian active learning.
- SMOCU approximates MOCU with a smooth strictly concave function.
- The strict concavity avoids the myopic behavior of *MOCU reduction*
- We prove the convergence of *SMOCU reduction*.