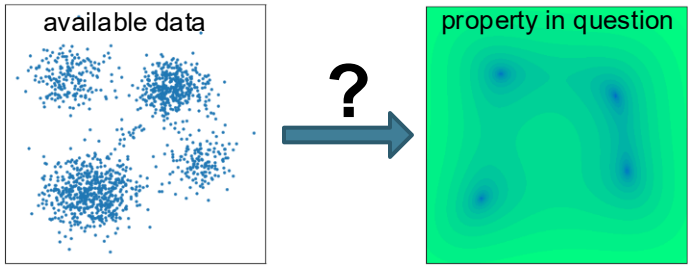
Liza Surzhikova (TU Brauschweig) | 06.11.2025

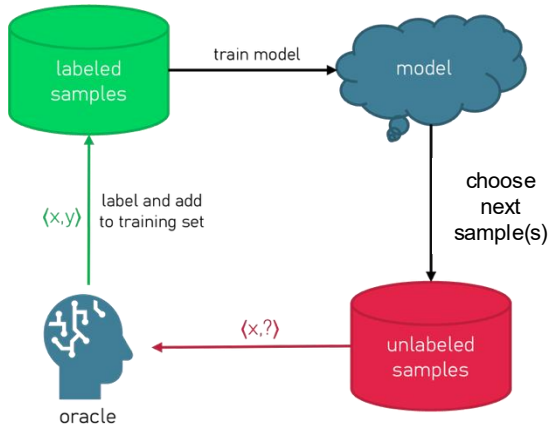
Active Learning

Assumption: Label generation most expensive step of training.
→ Learn property space with *as little labeled samples* as possible!

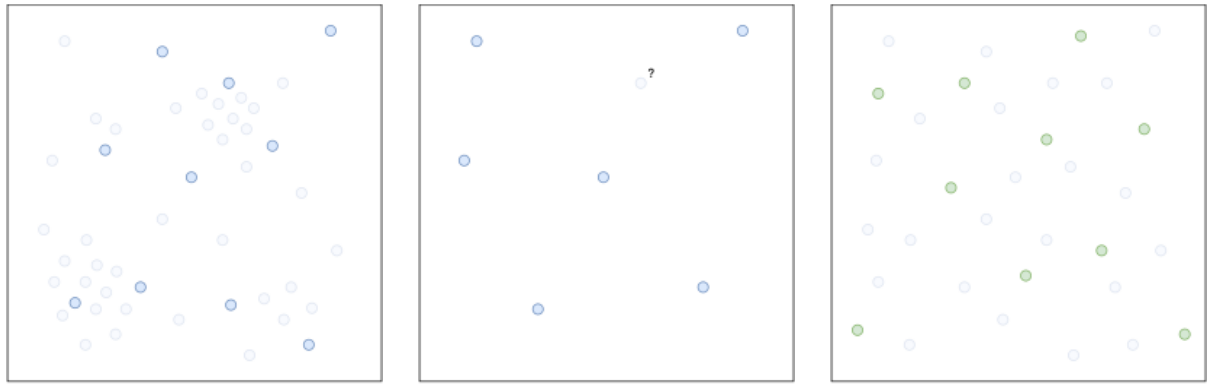


Active Learning

- iteratively develops training set with minimal labelling effort
- uses most informative available samples to avoid redundancy
- can pick one or batch of samples per iteration
- applicable to classification and regression problems

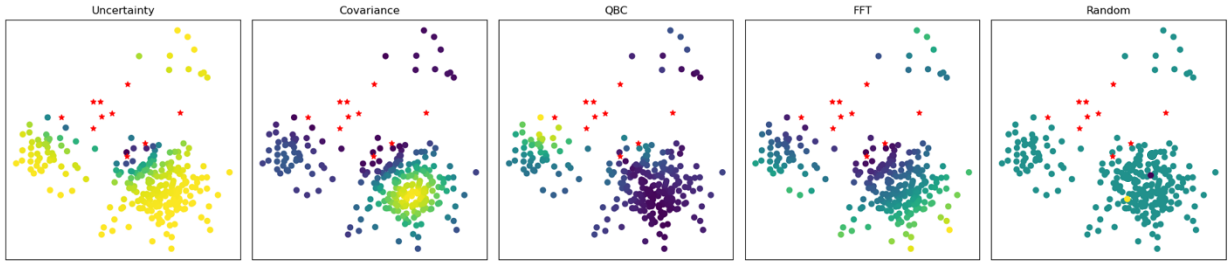


Pool-based setting | Stream-based setting | Synthetic setting

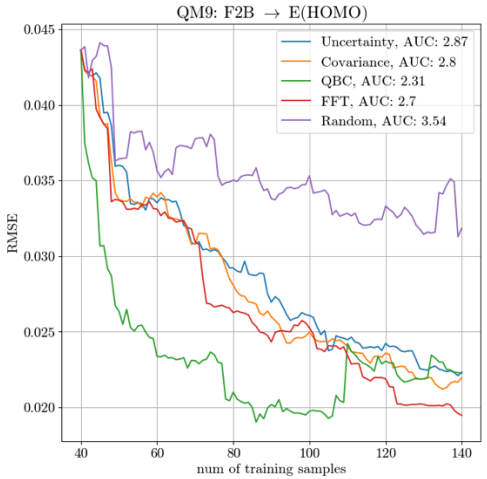


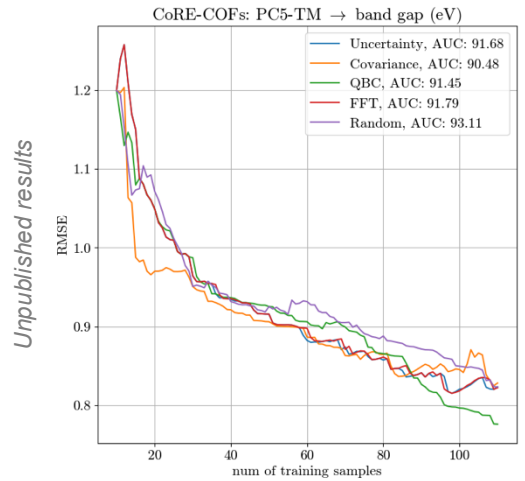
unlabeled sample | labeled sample | newly synthesized sample

Pool-based AL Strategies



Performance Examples





Applications in Chemistry

- Chemical space exploration
- Dataset reduction and/or construction
- Experiment prioritization

References
J. Comput. Sci. Technol. **2020**, *35*, 913-945.
Machine Learning Science and Technology **2025**, *6*, 025064.
J. Chem. Inf. Model. **2023**, *63*, *21*, 6515-6524.
J. Chem. Phys. **2018**, *148*, 241733.