Caterina Urban
ANTIQUE Research Team
Inria & École Normale Supérieure
caterina.urban@inria.fr

Robust Training of Support Vector Machines

SAVer (SVM Abstract Verifier, https://github.com/abstract-machine-learning/saver) [1] is an abstract interpretation-based method and tool for proving robustness or vulnerability properties of support vector machines (SVMs).

Goals. This project aims at designing a novel method for training robust (both linear and non-linear) SVMs based on abstract interpretation.

The main idea is to exploit an approximate worst-case adversarial loss function defined using SAVer. As a result, this novel training algorithm should provide SVM classifiers that are significantly more robust while retaining their accuracy with respect to a natural training. This work includes a first design phase for defining the new training algorithm, followed by an extensive experimental phase for assessing this new method.

Useful Prerequisites. The following skills would be helpful, but can also be learned during the project:

- Background in static analysis and abstract interpretation
- Familiarity with SVMs
- Experience with C++

Note. The project will be co-supervised by Prof. Francesco Ranzato (University of Padova, Italy, https://www.math.unipd.it/~ranzato/).

Contacts

- Caterina Urban
  caterina.urban@inria.fr

References

[1] Francesco Ranzato and Marco Zanella. Robustness verification of support vector machines. In Bor-Yuh Evan Chang, editor, *Static Analysis*, pages 271–295, Cham, 2019. Springer International Publishing.