



universität
wien

Fakultät für Mathematik
Oskar-Morgenstern-Platz 1
A-1090 Vienna
Austria

<https://owos.univie.ac.at>

ONE WORLD OPTIMIZATION SEMINAR

June 15th, 2020 @ 15:00 CEST (Central European Summer Time)

ALEXANDRE D'ASPREMONT

(École Normale Supérieure Paris)

Feature Selection & the Shapley - Folkman Theorem

Abstract. Due to its linear complexity, naive Bayes classification remains an attractive supervised learning method, especially in very large-scale settings. We propose a sparse version of naive Bayes, which can be used for feature selection. This leads to a combinatorial maximum-likelihood problem, for which we provide an exact solution in the case of binary data, or a bound in the multinomial case. We prove that our bound becomes tight as the marginal contribution of additional features decreases. Both binary and multinomial sparse models are solvable in time almost linear in problem size, representing a very small extra relative cost compared to the classical naive Bayes. Numerical experiments on text data show that the naive Bayes feature selection method is as statistically effective as state-of-the-art feature selection methods such as recursive feature elimination, l1-penalized logistic regression and LASSO, while being orders of magnitude faster. For a large data set, having more than with 1.6 million training points and about 12 million features, and with a non-optimized CPU implementation, our sparse naive Bayes model can be trained in less than 15 seconds.

The talk is based on joint work with Armin Askari and Laurent El Ghaoui that can be found at <https://arxiv.org/abs/1905.09884>.

The link of the zoom-room of the meeting and the corresponding password will be announced the day before the talk on the mailing list of the seminar, to which one can subscribe on <https://owos.univie.ac.at>.