

Machine Learning II: Assignment #2
14 performance points (max),
email: PDF+code to jan.nagler@gmail.com
due: Thu, April 18, 2024, A.D.

1. Kernels (7 points)

Based on the program developed in the lecture, implement a nonlinear Kernel classification based on SVM for polynomial kernels (poly) and for Gaussian Kernels, i.e., radial base functions (rbf). Apply the analysis to a database of your choice, using the python code as provided (or your own). Choose (and justify in a small text paragraph) two metrics to optimize for, such as accuracy, precision, recall, f1, f0.5, and f2 score, and determine which combo of hyperparameters maximizes each of the two scores you chose. It would be on you if you want to use GridSearch (as used in the provided code), RandomSearch, or other methods such as hyperopt. It is also on you to try other kernels of your choice. A 2-hyperparameter scan is sufficient. Of course, you are free to scan more parameters, if useful and interesting and doable for you. Do not forget to comment your choices and results in short take-home statements. Data can be found everywhere, e.g., at <https://archive.ics.uci.edu/ml/index.php>. No submissions based on cancer or wine data! Have fun. It will be.

2. Bomfunk MC Free Styler Regularization (7 points)

Study the uploaded notebook on regularization and improve it your way. For example, better visualize which features are important. Better visualize which features are correlated. Are certain heatmaps helpful? Do this at least for one other dataset of your choice. Do it your way. No submissions based on cancer or wine data!