

Machine Learning II: Assignment #2
14 performance points (max),
email: PDF+code to jan.nagler@gmail.com
due: Mon, May 11, 2020

1. Comments

Do not submit to my FS email address. Use my gmail address. Submissions with your full name in the filename only. If you team up, declare this with the 2 full names in the email to me and in the code in the first paragraph. Minimally comment your code. Attend the lecture on time. Thank you for your comments and feedback over email and Zoom meetings.

2. Kernels

Based on the program developed in the lecture, implement a nonlinear Kernel classification based on SVM for

(a)

(i) polynomial kernels (poly) and (ii) for Gaussian Kernels, i.e., radial base functions (rbf).

Apply the analysis to the breast cancer database, using the kernel python code as provided (or your own).

For this, consider accuracy, precision, recall, f1 and f2 score, and figure out what combo of hyperparameters (gamma) and degree (for polynomials) and penalty C maximizes the f2 score (see python code).

Which kernel works best? Plot the f2 score as a function of the hyperparameters (Your choice if you want to use 3d plots or 2d plots). Plot the confusion matrix for you best combo and the f2 score.

It is on you if you want to use GridSearch, RandomSearch, hyperopt, or just to try a few combinations of the hyperparameters. It is on you to also try other kernels of your choice.

(b, optional) Imagine you work for a company and you will have to communicate which method works best, for the given database. But you cannot maximize simultaneously all scores of the confusion matrix, think of which scores in particular are more important than others, given that you want to avoid false positives of malignant cancer. If you are not happy to focus on the f2 score, choose some other way and implement this.