University of Southern Denmark
IMADA

DM566: Data Mining and Machine Learning

Spring term 2022

Exercise 14

Exercise 14-1 Linear vs. non-linear classification

Use the SVM module of SKLearn to train a classifier on the following three datasets from the UCI Machine Learning Repository:

- Breast cancer diagnosis (https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29)
- Wine classification (https://archive.ics.uci.edu/ml/datasets/Wine)
- Yeast protein site classification (https://archive.ics.uci.edu/ml/datasets/Yeast)

Try different kernel types (e.g. linear, squared-exponential, polynomial) and different C values (e.g. 0.01, 0.1, 1, 10, 100). Plot the accuracy under these settings (For example, a with accuracy on the y-axis and $\log C$ on the x-axis for each fixed kernel type, and vice versa). Compare the plots for diffreent kernel types and comment on the effects of kernel type and C value on accuracy.