

## Homework 2: Linear Classification

Implement the pocket algorithm (linear classifier), with the following requirements:

- Initialize  $w$  using
  1. Method 1: the first data point
  2. Method 2: linear regression.
- Vectorize the computation when possible, and implement your own binary classifier (pocket algorithm) and linear regression routines.
- You may/should call the build-in function for pseudo-inverse.

Train and validate your implementation with 5 different sample sets (5 different  $D$ s in the learning flow) using the following dataset from sklearn:

- Breast cancer, 2 classes, 30 features, 569 data points

### To Submit:

Code your work with Python 3. You are encouraged to code your work with Jupyter Notebook.

You are supposed to submit both the well-documented .py python files (20pt) and the report. In the report, the following sections are required:

1. **Solution:** (10 pts) Provide the mathematical formulation of your algorithm, with different initialization mechanisms. The description should be independent of the programming language. (So this is not an explanation of your code)
2. **Training and validation:** (20 pts)
  - **Experiment:** Description of the setup of the experiment for both datasets, and evaluation of  $E_{in}(g)$  and  $E_{out}(g)$ .
  - **Result:** Show the performance plots using  $E_{out}(g)$  for  $t$  (number of iterations) for both initialization methods.
  - **Discussion:** Discuss the result: what did you observe? what do you learn from this observation?

Put all files together and submit a zipped file. Include a readme, explaining which problem(s) you have finished, so I know how to grade. Content in the readme file:

1. What did you finish?
2. What platform did you use (linux? Mac? windows?)
3. Resources that helped me.

Only one submission per team, and full names of the members should be included in the Comments box of the submission page on Blackboard.