Homework 2: Linear Classification

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

- Perform exploratory data analysis (EDA) and necessary data transformation (DT).
- Initialize w using 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 10 different sample sets (10 different Ds in the learning flow) with various N using the following dataset from sklearn:

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

To learn more about the dataset, visit

- https://scikit-learn.org/1.5/modules/generated/sklearn.datasets.load_breast_cancer.html

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/.ipynb python files (20pt) and the report. In the report, the following sections are required:

1. Solution: (20 pts) Provide the steps and results of EDA and DT. 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, and evaluation of $E_{in}(g)$ and $E_{out}(g)$.
    - Result: Show the performance plots of E_{in} and E_{out} for N (number of in-sample data points) for the datasets with and without DT, respectively.

- Discussion: What did you observe from the plots? What have you learned from the experiment?

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.