DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

LAB 1: Linear Classification

Machine Learning 2023
P. Zanuttigh, F. Chiariotti, M.Caligiuri, F. Lincetto, P. Talli

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

LAB 1: Linear Classification

Classification Task

Each training sample contains 3 features from a Telecom company in California

1. *Tenure in Months*
2. *Monthly Charge*
3. *Age of the customer*

Task: classify customer into 2 classes: "stayed" and "churned" (i.e., lost customer) based on the given features

Your Task

- You have to complete the jupyter notebook, tackling the classification problem
- FIRST THING TO DO: you need to put your name and ID number in the notebook
  - You can use the ID also as seed for random number generators
- The notebook has code to be completed: need to fill all required pieces
- You must write the answer to all the questions in the notebook (or you will lose points!)
- You should also place some text/comments (to explain choices or describe results)
- But do not change filenames, directory structure or the input data files, they will not be submitted

The Task

- Complete the jupyter notebook
  - i.e., write the code and answer to the questions
  - Place the questions' answers in the blue boxes
- Check that runs properly from the beginning with the provided data
  - use the `"restart kernel&run all"` command
- Save it as `surname_name_lab1.ipynb`
- Submit on elearning
- The homework is an individual task

Timeline

- Fri 3/11: Homework released
- Fri 10/11: Lab 1 (room Te and Ue)
- Fri 17/11: Delivery deadline
- The grade is +1 (to be added to the exam mark) if the homework is reasonably done
  - Lower score in case of very critical problems (e.g., missing parts or copied material)

Recall for LAB: Perceptron

Init: num_misclassified = -1: used to exit if there is no error

Normalize features for better performances

Input: training set $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)$
initialize $\mathbf{w}^{(1)} = (0, \ldots, 0)$;
for $t = 1, 2, \ldots$ **do**
  if $\exists i$ *s.t.* $y_i \langle \mathbf{w}^{(t)}, \mathbf{x}_i \rangle \leq 0$ **then** $\mathbf{w}^{(t+1)} \leftarrow \mathbf{w}^{(t)} + y_i \mathbf{x}_i$;
  else return $\mathbf{w}^{(t)}$;

Select at random

Need to select an error

"else" condition reached only for linealrly separable data

- Need stop condition
- Keep track of best solution (no guarantee that the last is the best)