

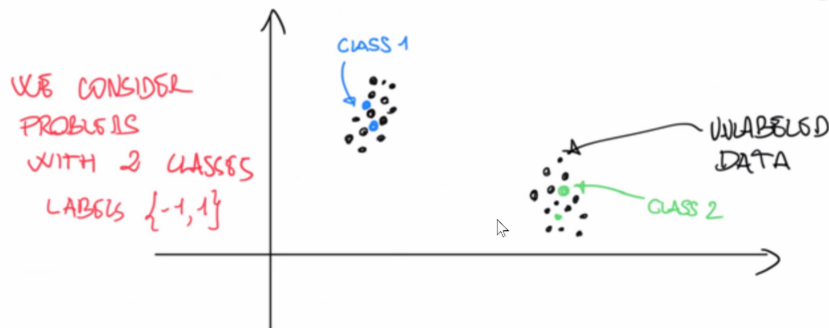
# Homework 1

## SEMI-SUPERVISED LEARNING

- WE HAVE  $\ell$  LABELED EXAMPLES  $(\bar{x}^i, \bar{y}^i)$   $i=1, \dots, \ell$
- WE HAVE  $u$  UNLABELED EXAMPLES  $x^j$   $j=1, \dots, u$
- GOAL: FIND  $y^i$ !

## IN REAL-WORLD PROBLEMS

- EASY TO GET DATA
- **HARD** TO GET LABELS  $\rightarrow$  HIGH NUMBER OF UNLABELED DATA



## PARADIGM SIMILAR FEATURES $\equiv$ SIMILAR LABELS

- DEFINE WEIGHTS  $w_{ij}$   $\rightarrow$  SIMILARITY BETWEEN
  - LABELED EXAMPLES  $i$
  - UNLABELED EXAMPLES  $j$
- DEFINE WEIGHTS  $\bar{w}_{ij}$   $\rightarrow$  SIMILARITY BETWEEN
  - UNLABELED EXAMPLES  $i, j$

## SQWS PROBLEM

$$(*) \min_{y \in \mathbb{R}^u} \underbrace{\sum_{i=1}^{\ell} \sum_{j=1}^u w_{ij} (y^j - \bar{y}^i)^2}_{\text{TERM 1}} + \underbrace{\frac{1}{2} \sum_{i=1}^u \sum_{j=1}^u \bar{w}_{ij} (y^i - y^j)^2}_{\text{TERM 2}}$$

$\bar{w}_{ij} = \bar{w}_{ji}$

- **TERM 1**: UNLABELED EXAMPLES SIMILAR LABELS TO LOSS LABELED ONES
- **TERM 2**: SIMILAR UNLABELED EXAMPLES GET SIMILAR LABELS

## HOW TO CHOOSE THE WEIGHTS?

Use some similarity measures based on features.

### HOMEWORK 1 (DEADLINE MAY 11TH)

① RANDOMLY GENERATE A SET OF POINTS IN 2D AND GIVE LABELS TO A SMALL SUBSET  $\ell$  OF THOSE POINTS

② CHOOSE A PROPER SIMILARITY MEASURE TO DEFINE  $w_{ij}$   $\bar{w}_{ij}$

③ CONSIDER THE PROBLEM (\*)

④ SOLVE PROBLEM (\*) WITH

(A) GRADIENT DESCENT

(B) BCD WITH RANDOMIZED RULES

(C) BCD WITH CYCLIC RULES

FOR (B)-(C) USE BLOCKS OF DIMENSION 1

⑤ CHOOSE A PUBLICLY AVAILABLE DATASET **TS** AND TEST THE METHODS ON THIS

⑥ ANALYZE ACCURACY VS COMPUTING (PLOTS)

⑦ DESCRIBE WHAT YOU DID ON A PDF FILE

(3 pages at most)

⑧ PUT FILES ON **GITHUB** FOLLOWING INSTRUCTIONS ON README

### FREE TIPS

• GRADIENT WRT  $y^j$

$$\nabla_{y^j} f(y) = 2 \sum_{i \in \ell} w_{ij} (y^j - \bar{y}^i) + 2 \sum_{i=1}^u \bar{w}_{ij} f(y^j - y^i)$$

$\downarrow$   
j<sup>th</sup> - COMPONENT  
OF THE GRADIENT