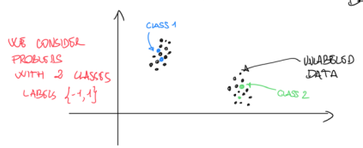
SEMISUPERVISED LEARNING

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

IN REAL-WORLD PROBLEMS

- EASY TO GET DATA
- HARD TO GET LABELS $\longrightarrow$ HUGE NUMBER OF UNLABELED DATA

WE CONSIDER PROBLEMS WITH 2 CLASSES LABELS $\{-1, 1\}$



PARADIGM SIMILAR FEATURES $\equiv$ SIMILAR LABELS

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

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

SOME PROBLEM

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

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

(*) HOW TO CHOOSE WEIGHTS?

- USE SOME SIMILARITY MEASURES BASED ON FEATURES!

HOMEWORK 1 (DEADLINE MAY 15 TH)

1. RANDOMLY GENERATE A SET OF POINTS IN 2D AND GIVE LABELS TO A SMALL SUBSET ℓ 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) BFGS WITH RANDOMIZED RULE
   - (C) BFGS WITH GS RULE

   FOR (B)-(C) USE BLOCKS OF DIMENSION 1
5. CHOOSE A PUBLICLY AVAILABLE DATASET TO AND TEST THE METHODS ON THIS
6. ANALYZE ACCURACY VS EPOCHS (PLOTS)
7. DESCRIBE WHAT YOU DID ON A PDF FILE
8. PUT FILES ON GITHUB FOLLOWING INSTRUCTIONS ON MOODLE

FREE TIPS

- GRADIENT WRT y^k

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