

project

笔记本: OFDS

创建时间: 2021/4/22 10:37

更新时间: 2022/1/27 13:10

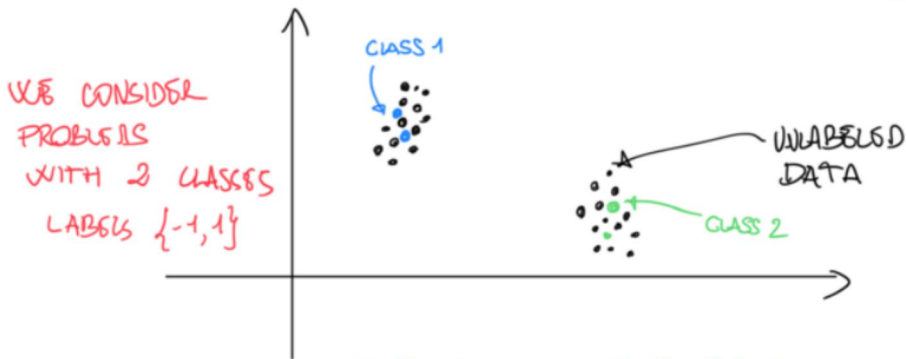
作者: 黄泽森

SEMI-SUPERVISED LEARNING

- WE HAVE ℓ 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

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

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

variable

let this term be as close to 0 as possible

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

⊗ HOW TO CHOOSE WEIGHTS?

similarity measure

- 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 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) BGD WITH RANDOMIZED RULES
 - (C) BGD WITH CYCLIC RULESFOR (B)-(C) USE BUCKS OF DIMENSION 1
- ⑤ CHOOSE A PUBLICLY AVAILABLE DATASET TS AND TEST THE METHODS ON THIS
accuracy VS CPUtime
- ⑥ ANALYZE ACCURACY VS COMPUTING (PLOTS)
- ⑦ DESCRIBE WHAT YOU DID ON A PDF FILE
- ⑧ PUT FILES ON GITHUB FOLLOWING INSTRUCTIONS ON README

FREE TIPS

gradient

- GRADIENT WRT y^j

$$\textcircled{*} \left[\nabla_{y^j} f(y) = 2 \sum_{i=1}^{\ell} w_{ij} (y^j - \bar{y}^i) + 2 \sum_{i=1}^u \bar{w}_{ij} (y^j - y^i) \right]$$