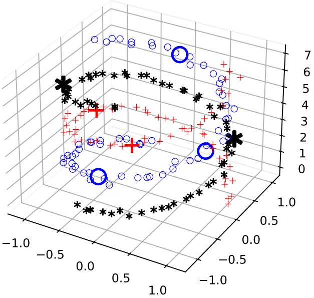
Label Propagation via Diffusion on a Graph

Julian Schmitt, Stefano Ferioli, Camilla Beneventano
École Polytechnique Fédérale de Lausanne, Switzerland

December 7, 2023

First concepts:

- Solve a classification problem with very small amount of labelled data
- Aim to capture information about the geometry of the data
- A graph provides information about the distance between points



Method

Setting: labeled data $(x_i, y_i)_{i=1}^{l}$ and
unlabeled data $(x_i)_{i=l+1}^{u}$ with $x_i \in \mathbb{R}^m$ and $y_i \in \{0,1\}$

1) Define an edge-weight matrix W for $G = (V, E)$ as

$$w_{ij} = exp\left(-\sum_{d=1}^{m} \frac{(x_{id} - x_{jd})^2}{\sigma_d^2}\right)$$

2) Find $f\colon V \to \mathbb{R}$ on G s.t. $f|_L = f_l$ via *Harmonic Energy Minimization* that is solving the problem

$$f = \arg\min_{f|_L = f_l} E(f) \quad \text{with} \quad E(f) := \frac{1}{2} \sum w_{ij}(f(i) - f(j))^2.$$

3) Label the data using f

Solution of the minimization problem

Set $D := \operatorname{diag}\left(\left(\sum_j w_{ij}\right)_i\right)$ and $P = D^{-1}W$.
The solution of the minimization problem is obtained by solving

$$\Delta f = 0$$

where Δ is the combinatorial Laplacian defined as $\Delta = D - W$.

This gives

$$f_u = (D_{uu} - W_{uu})^{-1} W_{ul} f_l = (I - P_{uu})^{-1} P_{ul} f_l.$$

with

$$W = \left[\begin{array}{c|c} W_{ll} & W_{lu} \\ \hline W_{ul} & W_{uu} \end{array}\right]$$

Remark: f_u is constant or $f_u \in (0,1)^u$ for maximum principle

Labelling the data using f

- Harmonic Threshold (Thresh):

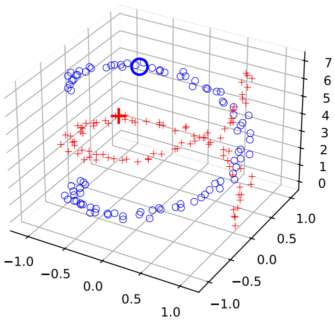
$$y_i = \begin{cases} 1 & \text{if } f(i) > \frac{1}{2}, \\ 0 & \text{if } f(i) \leq \frac{1}{2}. \end{cases} \quad \forall i \in U$$

- Class Mass Normalization (CMN):
  Let q and $1 - q$ the desirable proportions for classes 1 and 0 respectively. Then, set

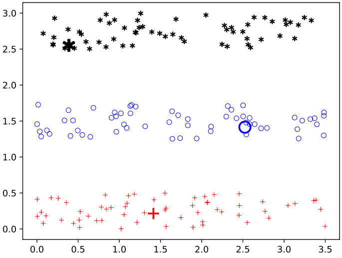
$$y_i = \begin{cases} 1 & \text{if } q \frac{f_u(i)}{\sum_i f_u(i)} > (1-q) \frac{1 - f_u(i)}{\sum_i (1 - f_u(i))}, \\ 0 & \text{otherwise.} \end{cases} \quad \forall i \in U$$

Remark: These methods can be generalized to multiclass problems using a 1-vs-rest approach

Harmonic energy minimization follows the structure of the data:



(a) 2 Spirals



(b) 3 Sets

Learning the Weight Matrix W

To retrieve optimal values for the weight parameters σ_d, we minimize the average label entropy $\min_\sigma H(f)$ with

$$H(f) = \frac{1}{u} \sum_{i=l+1}^{l+u} -f(i) \log f(i) - (1 - f(i)) \log (1 - f(i)).$$

Idea: H measures how *confident* the labelling is

Problem: H has a minimum in 0 as $\sigma_d \to 0$

Solution: Smooth the matrix P as $\tilde{P} = \varepsilon U + (1 - \varepsilon) P$ where $U_{ij} = \frac{1}{l+u}$

Expansion: incorporating External Classifiers

Goal: Combine an external classifier with harmonic energy minimization

Solution: Modify the Graph

- Introduce labeled "dongle" node i' for each unlabeled node i
- Transition probability from i' to i is η
- Transition probability of every other transition is multiplied by $1 - \eta$

The solution of this new optimization problem is given by

$$f_u = (I - (1-\eta)P_{uu})^{-1}((1-\eta)P_{ul}f_l + \eta h_u).$$

Classifiers for Comparison

- 1-NN
- Radial Basis Function Classifier (RBF):

$$y_i = \begin{cases} 1 & \text{if } (W_{ul} f_l)_i > (W_{ul}(1 - f_l))_i, \\ 0 & \text{otherwise.} \end{cases} \qquad \forall i \in U$$

- Nadaraya-Watson estimator with the Noyau Gaussien kernel

$$\hat{f}(x) = \sum_{i=1}^{l} \frac{s(x, x_i) y_i}{\sum_{j=1}^{l} s(x, x_j)} \qquad s(x, x_i) = \exp\left(-\frac{||x - x_i||^2}{h^2}\right)$$

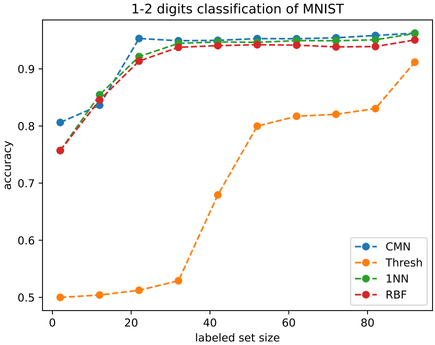
Experiments - MNIST

Apply the harmonic energy minimization to the MNIST dataset:

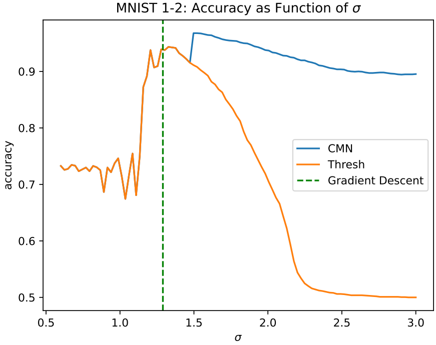
1. 1-2 binary classification
2. 10-digits multiclass classification
3. odd-even binary classification

Note: To simplify, we assumed there is only one single σ

MNIST - 1vs2

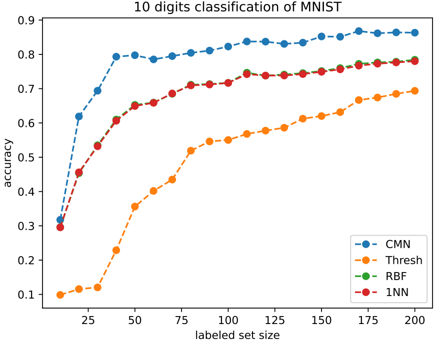


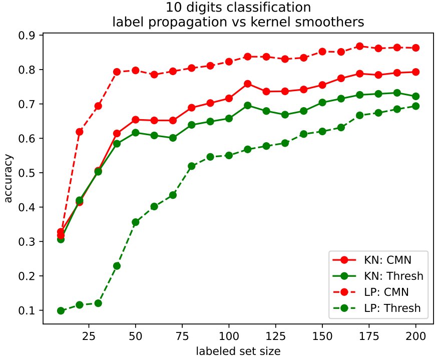
1-2 digits classification of MNIST
accuracy
labeled set size
0.5
0.6
0.7
0.8
0.9
0
20
40
60
80
CMN
Thresh
1NN
RBF



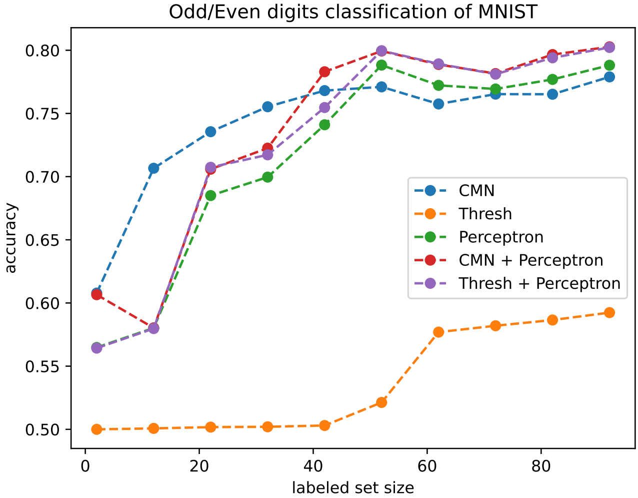
MNIST 1-2: Accuracy as Function of σ
accuracy
σ
0.5
0.6
0.7
0.8
0.9
0.5
1.0
1.5
2.0
2.5
3.0
CMN
Thresh
Gradient Descent

MNIST - 10 digits





MNIST - odd-vs-even



Conclusion

Open Questions / Possible Improvements:

- Optimize η when using external classifiers
- One-hot encoding instead of one-vs-all for multiclass problems