Graphical Semisupervised Learning

We should probably write something here

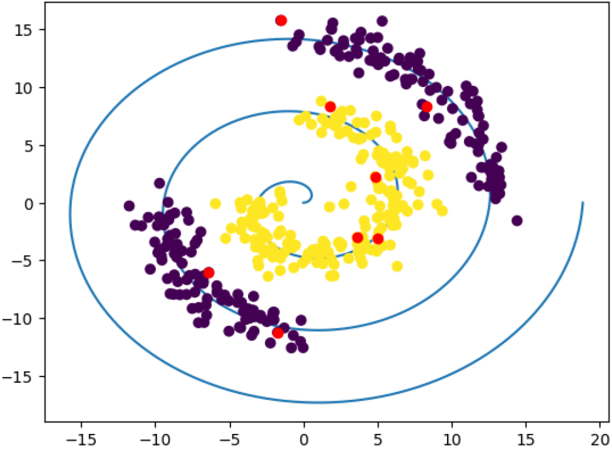
Rohin Gilman Alex Johnson Kaitlynn Lilly

May 22, 2023

UNIVERSITY of WASHINGTON | Applied Mathematics

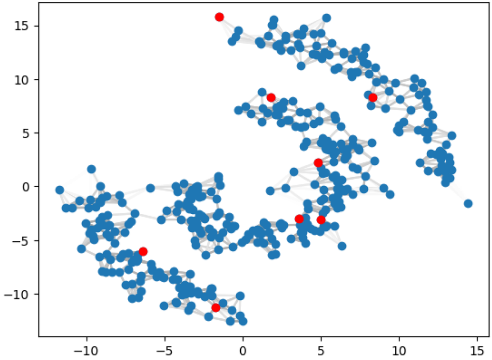
Introduction to Problem

- Goal in SSL is, given $\{x_i, y_i\}_{i=1}^{M}$ and $\{x_i\}_{i=M+1}^{N}$, to predict $y_i(x_i)$ for $i = M+1, \ldots, N$ where $M \ll N$.



GSSL

- We leverage spectral geometric properties of the graph Laplacian matrix L.
- First, we construct our graphs by KNN or Proximity graph construction.



Optimization Problem
Loss functions

- General optimization

$$\vec{f}^* = \underset{\vec{f}\in\mathbb{R}^m}{\arg\min}\, \mathcal{L}(\vec{f}; y) + \lambda \vec{f}^T C^{-1} \vec{f}$$

  where $\mathcal{L}$ is a loss function, either Probit or Regression loss in our case.

- Probit loss

$$\mathcal{L}(\vec{f}, y) = -\sum_{j=1}^{M} \log \Psi(\vec{f_j} y_j)$$

- Regression loss

$$\mathcal{L}(\vec{f}, y) = \sum_{j=1}^{M} \left(\vec{f_j} - y_j\right)^2$$

Optimization Problem
Matérn kernel-regularization

- General optimization

$$\vec{f}^* = \arg\min_{\vec{f}\in\mathbb{R}^m} \mathcal{L}(\vec{f}; y) + \lambda \vec{f}^T C^{-1} \vec{f}$$

where $C = (L + \tau^2 I)^{-\alpha}$, where L is the graph Laplacian.

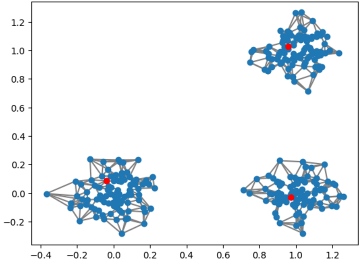
- In class, we showed C is a kernel matrix. In fact, C belongs to a class of kernels called the Matérn family that have the form

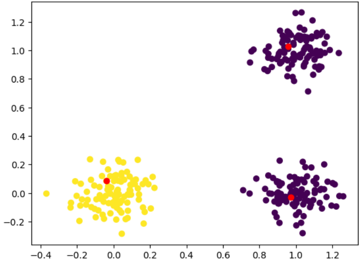
$$K(x, y) = \kappa(\|x - y\|), \kappa(t) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{t}{\gamma}\right) K_\nu\left(\sqrt{2\nu}\frac{t}{\gamma}\right)$$

where Γ is the Gamma function and K_ν is the modified Bessel function of the second kind.

Results
Three Cluster Case

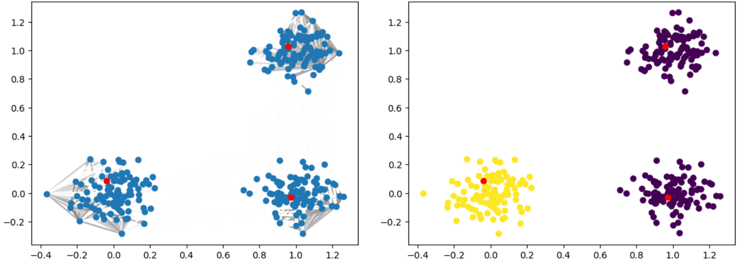
Using the K-Nearest Neighbors Graph with Uniform Weights





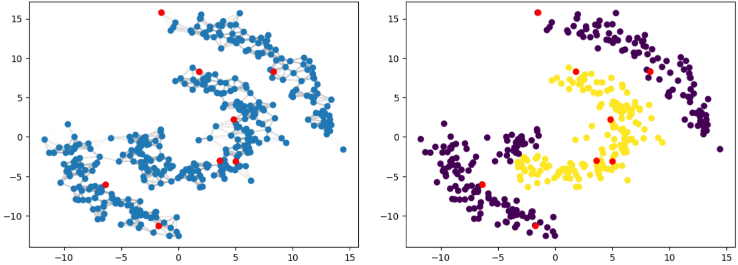
Loss Function	Classification Accuracy (over 50 trials)
Probit	100%
Regression	100%

Using the Proximity Graph with RBF Weights



Loss Function	Classification Accuracy (over 50 trials)
Probit	100.0%
Regression	100.0%

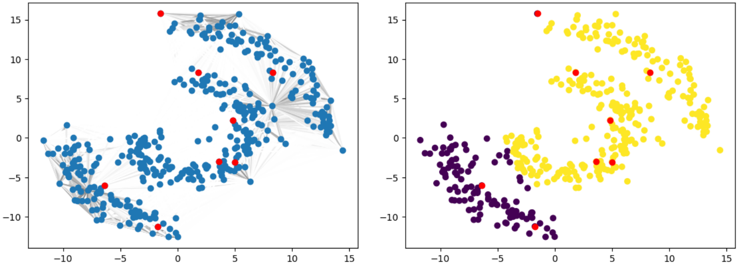
Using the K-Nearest Neighbors Graph with RBF Weights



Loss Function	Classification Accuracy (over 50 trials)
Probit	96.3%
Regression	95.8%

Results
Spiral Case

Using the Proximity Graph with RBF Weights



Loss Function	Classification Accuracy (on 1 trial)
Probit	73.0%
Regression	63.0%

Conclusions and Further Questions

- Conclusions
  - For well separated clusters, the method is effective, even without careful tuning
  - Not the case for more difficult scenarios
- Challenges Encountered
  - Tuning of paramters - especially for real data
  - Majority labelling
- Next Steps
  - Plan to apply this approach to MNIST to test on real data
  - Anticipate tuning challenges for more difficult pairs of numbers

Bibliography

Viacheslav Borovitskiy, Iskander Azangulov, Alexander Terenin, Peter Mostowsky, Marc Peter Deisenroth, and Nicolas Durrande.

Matérn Gaussian processes on graphs, April 2021.

Franca Hoffmann, Bamdad Hosseini, Zhi Ren, and Andrew M. Stuart.

Consistency of semi-supervised learning algorithms on graphs: Probit and one-hot methods, March 2020.

Daniel Sanz-Alonso and Ruiyi Yang.

The SPDE approach to Matérn fields: Graph representations, April 2021.