

Graph Convolution for Semi-Supervised Classification: Improved Linear Separability and Out-of-Distribution Generalization

Objective

- Study the effect of graph convolution on linear separability of a GMM
- Analyze generalization potential of the cross-entropy minimizer
- Conduct experiments in various settings to illustrate our results

Effect of Graph Convolution



*Supported by the Natural Sciences and Engineering Research Council of Canada (CRSNG), grant number [RGPIN-2019-04067, DGECR-2019-00147]. +Supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Cette recherches en sciences naturelles et en génie du Canada (CRSNG), grant number [RGPIN-2020-04597, DGECR-2020-00199].

Aseem Baranwal, Kimon Fountoulakis*, Aukosh Jagannath+

Separability Results



- Without the graph, no hyperpland $\|\mu - \nu\|_{2}$
- With graph convolution, this three

 $\|\mu - \nu\| = \mathcal{O}$

Without graph convolution, the lo

 $Loss \geq (2\log 2)$

• When the convolved data is separable, the loss is upper bounded

 $Loss(A, X) \leq C \exp (-$

Generalization

• For any new dataset A, X with different n, p, q, loss is bounded

 $Loss(A, X) \leq C \exp (-$

Loss increases with inter-class edge probability q (noisy graph)



The can separate a binary GMM if

$$p_2 = \mathcal{O}(\sigma)$$
The shold changes to
 $\mathcal{O}\left(\frac{\sigma}{\sqrt{ED}}\right) \rightarrow \text{Expected degree of a node}$
The second degree of a node
oss is lower bounded
 $\Phi\left(-\frac{\|\mu - \nu\|}{2\sigma}\right)$

$$-d\|\mu-\nu\|\left(\frac{p-q}{p+q}\right)\right)$$

$$-d\|\mu-\nu\|\left(\frac{p-q}{p+q}\right)\right)$$



Large q implies higher loss — graph convolution hurts linear separability of data

- linearly separable data

- models and deeper networks



Conclusions

Graph convolution can transform linearly inseparable data into

 Graph convolution can be disadvantageous if the intra-class edge probability is close to the inter-class edge probability

The learned classifier generalizes to out-of-distribution data

Future Work

• Extension to setting with multiple classes

• Properties of the optimal classifier for non-linearly separable data

Understanding effects of graph convolution for highly non-linear