Parent Item: Stochastic k-Neighborhood Selection for Supervised and Unsupervised Learning

Summary

Intro

- NCA is limited to k=1 but provides a useful way of metric learning
- This paper generalizes NCA to allow for any choice of k (equivalent to NCA at k=1)
- Also contribute techniques for computing expected accuracy and extension of SNE to k-neighborhood setting

Background

- They follow the linear version of NCA and also allow the same ability to set the dimension
- SNE is an unsupervised dimensionality reduction method (attempts to reproduce the structure of high-dim in a low-dim space)

Related Work

- Not aware of other NCA based work that extends k>1
- Outlines other metric learning methodologies

k-Neighborhood Components Analysis

- Two steps:
    1. define a probability distribution over selection of sets of k neighbors
    2. modify accuracy measure to reflect KNN procedure selects a label by majority vote

Efficient Computation

- Factor graphs are used to efficiently compute expected KNN accuracy
- Are able to efficiently compute partition funcs by using an alternative algo for computing marginal probabilities and partition funcs (use dynamic programming)
- Take advantage of the fact that any configuration having more than k neighbors is disallowed to improve efficiency
- This allows other choices for accuracy other than Majority
- For example, can get an accuracy measure that only rewards neighborhoods that have all neighbors of the target class

k-Stochastic Neighbors Embedding

- For unsupervised version they use SNE as the starting point and generalize using previous optimization calculations and methods

Experiments

- Use USPS handwritten digits and test on the two accuracy measures
- All trained models promote larger margin compared to majority trained which allow more dispersion in clusters and a smaller margin
- Since 1NCA is focused on immediate neighborhood, it can fall into many local optima and since Majority tends to forgive impure neighborhoods there are also more local optima

Discussion

- Provided a methodology for NCA learning with K>1
- can make an argument that using ALL has robustness and generalizes well since it forces margins between groups
- Can still extend to be nonlinear