Metric Learning for Clustering in Streaming Large-Scale Data

Parag Jain
Thesis advisor: Dr. Vineeth N Balasubramanian

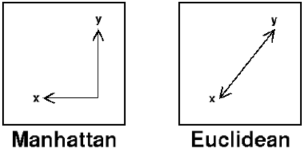
Indian Institute of Technology, Hyderabad

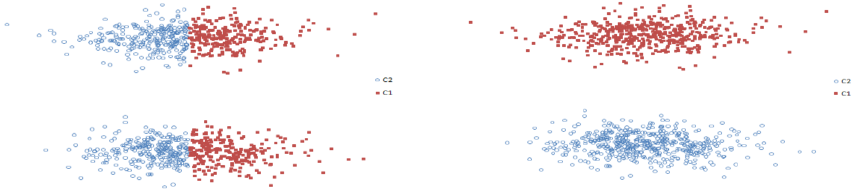
cs13m1008

January 2, 2016

Overview

Metric is important





Euclidean metric

Manhattan metric

k-means results

Image credit: www.improvedoutcomes.com

What is metric learning?

Metric learning

The goal of any metric learning method is to adapt some pair-wise metric function according to the problem of interest by learning from input training data.

General process

$\xrightarrow{\text{Data}}$ Metric learning $\xrightarrow{\text{Metric}}$ Metric based algorithm $\xrightarrow{\text{Model}}$ Prediction

Metric

Definition

A metric on a set X is a function (called the distance function).
$d : X \times X \to R$, *where R is a set of real numbers, and for all x,y,z in X following condition are satisfied:*

- $d(x, y) \geq 0$ *(non-negativity)*
- $d(x, y) = 0$ *if and only if* $x = y$ *(coincidence axiom)*
- $d(x, y) = d(y, x)$ *(symmetry)*
- $d(x, z) \leq d(x, y) + d(y, z)$ *(triangle inequality).*

If a function does not satisfy the second property but satisfies other three then it is called a **pseudometric**.

Background

Mahalanobis distance

Most of the metric learning methods in literature learns the metric of form,

$$d_M(x, x') = \sqrt{(x - x')^T M (x - x')}$$

which is Mahalanobis distance,where, $M = (A^{1/2})^T (A^{1/2})$ is a positive semi-definite matrix.

Commonly used constraints[1]

1. Similarity/dissimilarity constraints

$$d_A(x_i, x_j) \le u \quad (i,j) \quad \in S$$
$$d_A(x_i, x_j) \ge l \quad (i,j) \quad \in D$$

where, $(i,j) \in S$ for objects that are similar $(i,j) \in D$ for objects that are dissimilar.

[1] Kulis, Brian. "Metric learning: A survey." Foundations & Trends in Machine Learning 5.4 (2012)

Commonly used constraints[1]

1. Similarity/dissimilarity constraints

$$d_A(x_i, x_j) \leq u \quad (i,j) \quad \in S$$
$$d_A(x_i, x_j) \geq l \quad (i,j) \quad \in D$$

where, $(i,j) \in S$ for objects that are similar $(i,j) \in D$ for objects that are dissimilar.

2. Relative constraints
$R = \ (x_i, x_j, x_k) : x_i$ should be more similar to x_j than to x_k .:

$$d_A(x_i, x_j) < d_A(x_i, x_k) - m$$

Where m is margin, generally m is chosen to be 1.

[1] Kulis, Brian. "Metric learning: A survey." Foundations & Trends in Machine Learning 5.4 (2012)

Motivation from Ocean Data Analysis

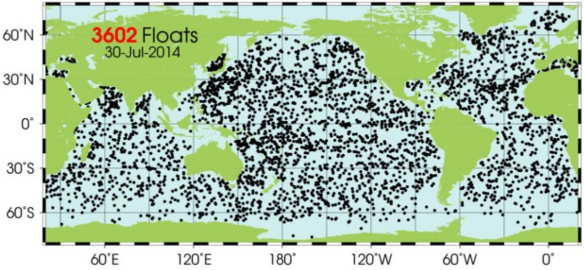


Figure: ARGO around the world

Figure: A float

Image credit: www.argo.ucsd.edu

Motivation from Ocean Data Analysis

This problem was brought up by scientists at INCOIS[2].

$$\left\{\begin{array}{c}\text{temperature}\\ \text{salinity}\\ \text{oxygen}\\ \text{nitrate}\\ \text{phosphate}\\ \text{silicate}\\ .\\ .\end{array}\right\}$$

Data format

- Clustering is used by scientist and experts to visualize and detect trends in ocean behaviour.

[2] Indian National Centre for Ocean Information Services

Motivation from Ocean Data Analysis

This problem was brought up by scientists at INCOIS[2].

$$\left\{ \begin{array}{c} \text{temperature} \\ \text{salinity} \\ \text{oxygen} \\ \text{nitrate} \\ \text{phosphate} \\ \text{silicate} \\ . \\ . \end{array} \right\}$$

Data format

- Clustering is used by scientist and experts to visualize and detect trends in ocean behaviour.
- How to correctly cluster the enormous amount of data? How to choose a distance metric?

[2] Indian National Centre for Ocean Information Services

Motivation from Ocean Data Analysis

This problem was brought up by scientists at INCOIS[2].

$$\left\{\begin{array}{c}\text{temperature}\\ \text{salinity}\\ \text{oxygen}\\ \text{nitrate}\\ \text{phosphate}\\ \text{silicate}\\ .\\ .\end{array}\right\}$$

Data format

- Clustering is used by scientist and experts to visualize and detect trends in ocean behaviour.
- How to correctly cluster the enormous amount of data? How to choose a distance metric?
- High dimensional data makes it difficult to manually derive a metric.

[2] Indian National Centre for Ocean Information Services

Motivation from Ocean Data Analysis

This problem was brought up by scientists at INCOIS[2].

$$\left\{ \begin{array}{c} \text{temperature} \\ \text{salinity} \\ \text{oxygen} \\ \text{nitrate} \\ \text{phosphate} \\ \text{silicate} \\ . \\ . \end{array} \right\}$$

Data format

- Clustering is used by scientist and experts to visualize and detect trends in ocean behaviour.
- How to correctly cluster the enormous amount of data? How to choose a distance metric?
- High dimensional data makes it difficult to manually derive a metric.
- Continuous change in world climate makes ocean data ever evolving, which means updating the metric is a challenge.

[2] Indian National Centre for Ocean Information Services

Challenges in Metric Learning

- How to learn a distance metric in a supervised setting?
- How to learn a metric in an unsupervised setting?
- How to metric a update as new data arrives?
- How to update metric online by intelligently selecting points from new data to query user?

Challenges in Metric Learning

- How to learn a distance metric in a supervised setting?(Many recent work available.)
- How to learn a metric in an unsupervised setting?We contribute an unsupervised method to learn a metric.

Challenges in Metric Learning

- How to learn a distance metric in a supervised setting?(Many recent work available.)
- How to learn a metric in an unsupervised setting?We contribute an unsupervised method to learn a metric.
- How to metric a update as new data arrives? Online methods available for supervised setting.We propose an incremental unsupervised method.

Challenges in Metric Learning

- How to learn a distance metric in a supervised setting?(Many recent work available.)
- How to learn a metric in an unsupervised setting?We contribute an unsupervised method to learn a metric.
- How to metric a update as new data arrives? Online methods available for supervised setting.We propose an incremental unsupervised method.
- How to update metric online by intelligently selecting points from new data to query user? Recent work for classification problems.We propose method for clustering setup.

Unsupervised Metric Learning using Low Dimensional Embedding

This approach combines Laplacian Eigenmaps with Information Theoretic Metric Learning to form an unsupervised metric learning approach.
Learns an unsupervised distance metric function.

Unsupervised Metric Learning using Low Dimensional Embedding

This approach combines Laplacian Eigenmaps with Information Theoretic Metric Learning to form an unsupervised metric learning approach.
Learns an unsupervised distance metric function.

Incremental Diffusion Maps

In this method we extended Diffusion Maps for incremental framework, using incremental SVD.
Learns all pairwise distance incrementally and thereby can be used for new data.

Unsupervised Metric Learning using Low Dimensional Embedding

This approach combines Laplacian Eigenmaps with Information Theoretic Metric Learning to form an unsupervised metric learning approach.
Learns an unsupervised distance metric function.

Incremental Diffusion Maps

In this method we extended Diffusion Maps for incremental framework, using incremental SVD.
Learns all pairwise distance incrementally and thereby can be used for new data.

Online Active Metric Learning for Clustering

We propose a two stage approach for online active metric learning based on active pairwise data selection.
Selects important points to query and apply online metric update.

Metric Learning Review

Some popular metric learning methods from survey[3]:

Type	Method
Supervised learning	MMC , Schultz and Joachims LMNN, ITML, POLA
Unsupervised learning	SSO, CPCM[4]
Online metric learning	POLA, LEGO, MDML
Manifold based	Diffusion Maps[5], LE, LLE, MDS

[3] Kulis, Brian. "Metric learning: A survey." Foundations & Trends in Machine Learning 5.4 (2012)

[4] Gupta, Abhishek A. "Unsupervised distance metric learning using predictability". Technical Reports (CIS).(2008)

[5] Coifman, R.R.; S. Lafon. "Diffusion maps". Applied and Computational Harmonic Analysis 21: 530. (2006)

Table of Content

Unsupervised Metric Learning using Low Dimensional Embedding

- We combine Laplacian eigenmaps[6] and Information Theoretic Metric Learning[7](ITML) to form Unsupervised Metric Learning method.

[6] Belkin, Mikhail. "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering." NIPS.2001.

[7] Davis, Jason V., et al. "Information-theoretic metric learning." ICML 2007.

Unsupervised Metric Learning using Low Dimensional Embedding

- We combine Laplacian eigenmaps[6]and Information Theoretic Metric Learning[7](ITML) to form Unsupervised Metric Learning method.
- Laplacian Eigenmaps
  - Projects data into low dimensional space which is euclidean in natures.
  - Extract similar & dissimilar pairs from low dimensional space.

[6]Belkin, Mikhail. "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering." NIPS.2001.

[7]Davis, Jason V., et al. "Information-theoretic metric learning." ICML 2007.

Unsupervised Metric Learning using Low Dimensional Embedding

- We combine Laplacian eigenmaps[6]and Information Theoretic Metric Learning[7](ITML) to form Unsupervised Metric Learning method.
- Laplacian Eigenmaps
  - Projects data into low dimensional space which is euclidean in natures.
  - Extract similar & dissimilar pairs from low dimensional space.
- Learn metric using ITML by using similar & dissimilar pair as constraints.

[6]Belkin, Mikhail. "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering." NIPS.2001.

[7]Davis, Jason V., et al. "Information-theoretic metric learning." ICML 2007.

Laplacian Eigenmaps

- Given set of data points $\mathcal{X} \in \mathcal{R}^N$, find an embedding in m dimensional space where $m < N$ such that the local geometry is optimally preserved.

$$x_i \in \mathcal{R}^N \mapsto y_i \in \mathcal{R}^m$$

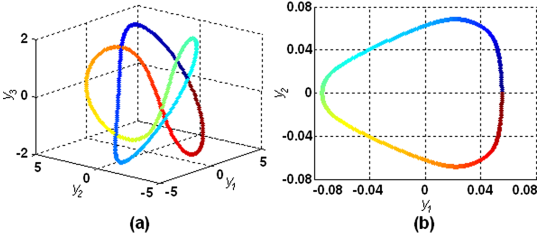


Image credit: electronicimaging.spiedigitallibrary.org

ITML: Information-Theoretic Metric Learning

- Main idea is to use a regularizer that automatically enforces PSD constraints on $M_{d \times d}$

LogDet formulation

$$\begin{aligned} \min_{A} \quad & D_{ld}(A, A_0) \\ \text{s.t.} \quad & d_A(x_i, x_j) \leq u \;\; \forall (x_i, x_j) \in S \\ & d_A(x_i, x_j) \geq l \;\; \forall (x_i, x_j) \in D \\ & A \succeq 0. \end{aligned}$$

$$D_{ld}(A, A_0) = tr(AA_0^{-1}) - logdet(AA_0^{-1}) - d$$

What are the limitations?

- Laplacian Eigenmaps gives you an embedding but no distance function is learned which you can use for out of sample data.

What are the limitations?

- Laplacian Eigenmaps gives you an embedding but no distance function is learned which you can use for out of sample data.
- ITML is a supervised method.

Algorithm

Input:
$X \in N \times k$, is input data in k dimensional space
t_s: threshold for similarity
t_d: threshold for dissimilarity
ϵ, m: parameters for laplacian eigenmaps algorithm, $m < k$
Output: Learned metric $A_{k \times k}$
Steps:

1. Construct low dimensional embedding:
   $\mathcal{E} = \text{laplacianEigenmaps}(X, \epsilon, m)$
2. Construct similarity and dissimilarity pairs:
   for each pair $(x_i, x_j) \in \mathcal{E}$:
   $p = ||x_i - x_j||^2$
   if $p \leq t_s$ then $S \leftarrow S \cup (x_i, x_j)$
   if $p \geq t_d$ then $D \leftarrow D \cup (x_i, x_j)$
3. Apply ITML procedure to learn metric:
   $A = \text{itml}(X, S, D)$

Advantages of proposed method

- Unsupervised learning, with applicability to out of sample points.

Advantages of proposed method

- Unsupervised learning, with applicability to out of sample points.
- Our method is flexible, we can decide how many constraints we want to pass to ITML.

Advantages of proposed method

- Unsupervised learning, with applicability to out of sample points.
- Our method is flexible, we can decide how many constraints we want to pass to ITML.
- Threshold parameter can be set easily by analysing distribution of pairwise distances in embedding space.

Advantages of proposed method

- Unsupervised learning, with applicability to out of sample points.
- Our method is flexible, we can decide how many constraints we want to pass to ITML.
- Threshold parameter can be set easily by analysing distribution of pairwise distances in embedding space.
- We can choose to project only new data to low dimensional space to get new similar dissimilar pairs.

Results

- Experimental setup: All results are average over 5 runs, with 80% randomly selected points used for training and 20% for testing. We have used k-NN for classification. All datasets are standard UCI[8]datasets.
- To best of our knowledge there is no method that does exactly what we tried so we compare our method with Euclidean distance.

Dataset	**Proposed method**	**Euclidean + ITML**
Letter recognition	95.25	93.75
Iris	83.3	76.6
Scale	84.7	82.6
Yeast	60.7	59.4
Wine	75.9	74.1

[9]Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml].

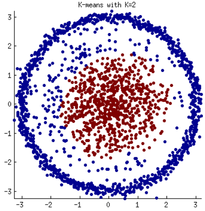
Incremental Diffusion Maps

$$[D^{(t)}(x_i, x_j)]^2 = \sum_{q \in \Omega} \frac{(p_{iq}^{(t)} - p_{jq}^{(t)})^2}{\varphi(x_q)^{(0)}}$$

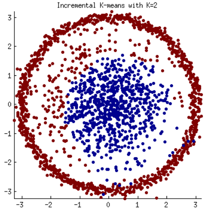
$$P = \begin{bmatrix} \psi_1(x_1) & \psi_1(x_2) & \cdots & \psi_1(x_{N_w}) \\ \psi_2(x_1) & \psi_2(x_2) & & \psi_2(x_{N_w}) \\ \vdots & \vdots & \vdots & \vdots \\ \psi_\alpha(x_1) & \psi_\alpha(x_2) & & \psi_\alpha(x_{N_w}) \\ \vdots & \vdots & \vdots & \vdots \\ \psi_{N_w}(x_1) & \psi_{N_w}(x_2) & \cdots & \psi_{N_w}(x_{N_w}) \end{bmatrix} \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_{N_w} \end{bmatrix} \begin{bmatrix} \vec{\phi}_1^T \\ \vec{\phi}_1^T \\ \vdots \\ \vec{\phi}_{N_w}^T \end{bmatrix}$$

$$[D^{(t)}(x_i, x_j)]^2 \cong \sum_{l=1}^{\alpha} (\lambda_l^t)^2 (\psi_l(x_i) - \psi_l(x_j))^2$$

We used Incremental SVD[9] to approximate diffusion distance for new points.



Batch mode. **Run time: 27.5 sec**



Incremental mode. **Run time 6.4 sec**

[9] 'On updating problems in latent semantic indexing,' SIAM (1999)

Online Active Metric Learning for Clustering

Introduction

- Unsupervised learning is ambitious and difficult, at the same time getting all the labels in a large dataset for training supervised algorithm is costly

Online Active Metric Learning for Clustering

Introduction

- Unsupervised learning is ambitious and difficult, at the same time getting all the labels in a large dataset for training supervised algorithm is costly
- More practical way is to get a few labels or hints from an expert to guide our algorithm

Online Active Metric Learning for Clustering

Introduction

- We propose a heuristic based approach to intelligently select a pair of points from the data which are queried to the user for similarity or dissimilarity information.

Online Active Metric Learning for Clustering

Introduction

- We propose a heuristic based approach to intelligently select a pair of points from the data which are queried to the user for similarity or dissimilarity information.
- Why pairs?
  - It is easy to say whether a pair of points are similar or dissimilar than providing actual labels.
  - It fits in clustering framework.

Pseudo-Metric Online Learning Algorithm(POLA)[10]

- Let $\mathcal{X}$ denotes the feature space. Given intial matrix threshold pair (A_τ, b_τ). POLA updates matrix threshold pair to $(A_{\tau+1}, b_{\tau+1})$, where metric is of form,

$$d_A(x, x') = \sqrt{(x - x')' A (x - x')}$$

[10] Shalev-Shwartz, Shai, Yoram Singer, and Andrew Y. Ng. "Online and batch learning of pseudo-metrics." Proceedings of the twenty-first international conference on Machine learning. ACM, 2004.

Pseudo-Metric Online Learning Algorithm(POLA)[10]

- Let $\mathcal{X}$ denotes the feature space. Given intial matrix threshold pair (A_τ, b_τ). POLA updates matrix threshold pair to $(A_{\tau+1}, b_{\tau+1})$, where metric is of form,

$$d_A(x, x') = \sqrt{(x - x')' A (x - x')}$$

- Input to algorithm is a pair of instance and corresponding label. $z = (x, x', y) \in (\mathcal{X} \times \mathcal{X} \times +1, -1)$, where $y = +1$ if pair (x, x') are similar otherwise $y = -1$.

[10]Shalev-Shwartz, Shai, Yoram Singer, and Andrew Y. Ng. "Online and batch learning of pseudo-metrics." Proceedings of the twenty-first international conference on Machine learning. ACM, 2004.

Pseudo-Metric Online Learning Algorithm(POLA)[10]

- Let $\mathcal{X}$ denotes the feature space. Given intial matrix threshold pair (A_τ, b_τ). POLA updates matrix threshold pair to $(A_{\tau+1}, b_{\tau+1})$, where metric is of form,

$$d_A(x, x') = \sqrt{(x - x')' A (x - x')}$$

- Input to algorithm is a pair of instance and corresponding label. $z = (x, x', y) \in (\mathcal{X} \times \mathcal{X} \times +1, -1)$, where $y = +1$ if pair (x, x') are similar otherwise $y = -1$.
- Given current A & b. Prediction loss is defined as,

$$l_\tau(A, b) = max\{0, y_\tau(d_A(x, x')^2 - b) + 1\}$$

[10] Shalev-Shwartz, Shai, Yoram Singer, and Andrew Y. Ng. "Online and batch learning of pseudo-metrics." Proceedings of the twenty-first international conference on Machine learning. ACM, 2004.

Pseudo-Metric Online Learning Algorithm(POLA)[10]

- Let $\mathcal{X}$ denotes the feature space. Given intial matrix threshold pair (A_τ, b_τ). POLA updates matrix threshold pair to $(A_{\tau+1}, b_{\tau+1})$, where metric is of form,

$$d_A(x, x') = \sqrt{(x - x')' A (x - x')}$$

- Input to algorithm is a pair of instance and corresponding label. $z = (x, x', y) \in (\mathcal{X} \times \mathcal{X} \times +1, -1)$, where $y = +1$ if pair (x, x') are similar otherwise $y = -1$.
- Given current A & b. Prediction loss is defined as,

$$l_\tau(A, b) = max\{0, y_\tau(d_A(x, x')^2 - b) + 1\}$$

- $A_\tau \succeq 0$ and $b_\tau \geq 1$

[10] Shalev-Shwartz, Shai, Yoram Singer, and Andrew Y. Ng. "Online and batch learning of pseudo-metrics." Proceedings of the twenty-first international conference on Machine learning. ACM, 2004.

Update Procedure

- On receiving new input, algorithm updates matrix threshold pair (A_τ, b_τ) in two steps.

Let

$$\mathcal{C}_L = \{(A, b) \in \mathbb{R}^{n^2+1} : L_\tau(A, b) = 0\}$$

and

$$\mathcal{C}_A = \{(A, b) \in \mathbb{R}^{n^2+1} : A \succeq 0, b \geq 1\}$$

This is the Lossfree half-space ($\mathcal{C}_L$) and the Admissible cone ($\mathcal{C}_A$).

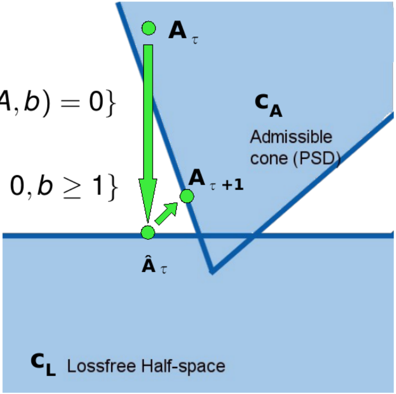


Image credit: cseweb.ucsd.edu elkan/254spring05

Selecting new pairs

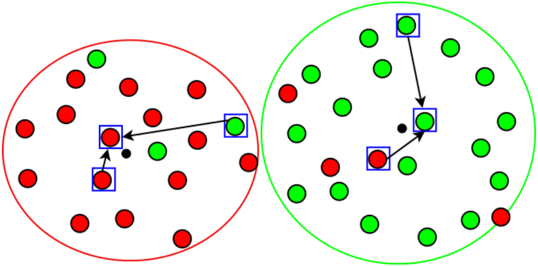


Figure: Illustration for selecting pairs

Algorithm 1 Online active metric learning

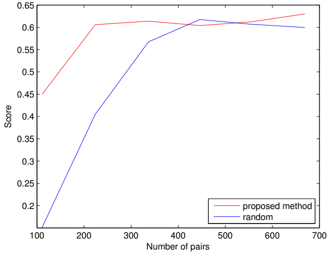
```
procedure OAML
    Input: number of cluster k
    N number of iterations
    X_{n×d} is input feature matrix
    r is ratio of near-center & boundary points to select
    p is number of pairs to select in each iteration
    α is alpha-trimming parameter
    Output: A learned metric
    Cluster belongingness matrix C_{n×k}
    Initialization: A ← I; S ← ; D ←
        for i = 1 to N do
    Ĉ ← kmeans(X, k, A)
    [S, D] ← SELECTPAIRS(X, C, p, S, D, A, r, α)
    Update matrix A based on new pairs using POLA, A ← pola(A, S, D)
        end for
end procedure
```

- Tested on Magic and Letter recognition datasets.
- To validate clustering Silhouette measure is applied which is given as $S = \frac{\sum_i s(i)}{n}$,

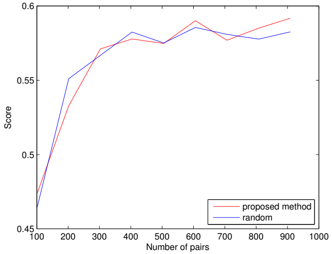
$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

  $a(i)$ is average distance of point x_i with points in the same cluster as of x_i and $b(i)$ is lowest average distance to points which are not in same cluster as x_i.
- To best of our knowledge no other method learn online metric using pairwise constraints for a clustering setup so we compare our method against randomly selected pairs.

Results



(a) Silhouette score for Letter recognition dataset



(b) Silhouette score for Magic dataset

Figure: Results of proposed method.

- All results are average over 5 runs.
- Label complexity of our method is less than randomly selecting pairs.

Conclusion

- There are has been a lot of work in supervised metric learning but unsupervised metric learning is still a challenge.
- Unsupervised metric learning can be difficult to address, we can using active learning methods to select informative points which is cost effective than a fully supervised setting.

Possible Extensions

- Our proposed OAML algorithm is heuristic based which has limitations, this can be improved by using a better approach with theoretical guarantees.

References I

Mikhail Belkin and Partha Niyogi.
Laplacian eigenmaps and spectral techniques for embedding and clustering.
In NIPS, volume 14, pages 585–591, 2001.

Aurélien Bellet, Amaury Habrard, and Marc Sebban.
A survey on metric learning for feature vectors and structured data.
arXiv preprint arXiv:1306.6709, 2013.

Derya Birant and Alp Kut.
St-dbscan: An algorithm for clustering spatial–temporal data.
Data & Knowledge Engineering, 60(1):208–221, 2007.

Xilun Chen and K Selcuk Candan.
Lwi-svd: low-rank, windowed, incremental singular value decompositions on time-evolving data sets.
In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 987–996. ACM, 2014.

Ronald R Coifman and Stéphane Lafon.
Diffusion maps.
Applied and computational harmonic analysis, 21(1):5–30, 2006.

Jason V Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S Dhillon.
Information-theoretic metric learning.
In Proceedings of the 24th international conference on Machine learning, pages 209–216. ACM, 2007.

References II

Sandra Ebert, Mario Fritz, and Bernt Schiele.
Active metric learning for object recognition.
Springer, 2012.

Ungar Lyle H. Gupta Abhishek A., Foster Dean P.
Unsupervised distance metric learning using predictability.
ScholarlyCommons Technical Reports (CIS), 2008.

Peng Jia, Junsong Yin, Xinsheng Huang, and Dewen Hu.
Incremental laplacian eigenmaps by preserving adjacent information between data points.
Pattern Recognition Letters, 30(16):1457–1463, 2009.

Jiayan Jiang, Bo Wang, and Zhuowen Tu.
Unsupervised metric learning by self-smoothing operator.
In Computer Vision (ICCV), 2011 IEEE International Conference on, pages 794–801. IEEE, 2011.

Brian Kulis.
Metric learning: A survey.
Foundations & Trends in Machine Learning, 5(4):287–364, 2012.

Gautam Kunapuli and Jude Shavlik.
Mirror descent for metric learning: a unified approach.
In Machine Learning and Knowledge Discovery in Databases, pages 859–874. Springer, 2012.

References III

Joshua M Lewis, Pincelli M Hull, Kilian Q Weinberger, and Lawrence K Saul.
Mapping uncharted waters: Exploratory analysis, visualization, and clustering of oceanographic data.
In Machine Learning and Applications, 2008. ICMLA'08. Seventh International Conference on, pages 388–395. IEEE, 2008.

M. Lichman.
UCI machine learning repository, 2013.

Sam T Roweis and Lawrence K Saul.
Nonlinear dimensionality reduction by locally linear embedding.
Science, 290(5500):2323–2326, 2000.

Shai Shalev-Shwartz, Yoram Singer, and Andrew Y Ng.
Online and batch learning of pseudo-metrics.
In Proceedings of the twenty-first international conference on Machine learning, page 94. ACM, 2004.

Sergios Theodoridis and Konstantinos Koutroumbas.
Pattern Recognition, Fourth Edition.
Academic Press, 4th edition, 2008.

Fei Wang and Jimeng Sun.
Distance metric learning in data mining.
SIAM SDM. Tutorials, 2012.

Fei Wang, Jimeng Sun, Tao Li, and Nikos Anerousis.
Two heads better than one: Metric+ active learning and its applications for it service classification.
In Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on, pages 1022–1027. IEEE, 2009.

References IV

Kilian Q Weinberger and Lawrence K Saul.
Distance metric learning for large margin nearest neighbor classification.
The Journal of Machine Learning Research, 10:207–244, 2009.

Eric P Xing, Michael I Jordan, Stuart Russell, and Andrew Y Ng.
Distance metric learning with application to clustering with side-information.
In Advances in neural information processing systems, pages 505–512, 2002.

Hongyuan Zha and Horst D Simon.
On updating problems in latent semantic indexing.
SIAM Journal on Scientific Computing, 21(2):782–791, 1999.

Thank you.