Classification using Minimax Distance

Madhav yadav

Indian Institute of Technology Hyderabad

Under the supervision of Dr.Balasubramaniam Jayaram

June 6, 2023

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Motivation for this work

Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Classification with Minimax Distance Measures

Morteza Haghir Chehreghani
Xerox Research Centre Europe - XRCE
6 chemin de Maupertuis
38240 Meylan, France
morteza.chehreghani@xrce.xerox.com

Decoding the title

Classification

Decoding the title

Classification

- A supervised ML process
  - categorizing a given set of input data into classes,
  - based on one or more variables.

Decoding the title

Classification

- A supervised ML process
  - categorizing a given set of input data into classes,
  - based on one or more variables.
- Real world data are very complex.
  - Basic distance functions fail to capture underlying patterns.

Minimax distance

- Each data point has two characteristics:
  - The vector
  - The Label.
- A dataset is represented in the form of graph $G(O, D)$

Minimax distance

- Each data point has two characteristics:
  - The vector
  - The Label.
- A dataset is represented in the form of graph $G(O, D)$
  - O is the set of objects representing data points,
  - D is the set of edges.

$d_{ij}-$ pairwise distance between nodes i and j.

Minimax distance

d_{ij} needs to satisfy the following conditions:

- $d_{ii} = 0$
- $d_{ij} \geq 0$
- $d_{ij} = d_{ji}$

Minimax distance

d_{ij} needs to satisfy the following conditions:

- $d_{ii} = 0$
- $d_{ij} \geq 0$
- $d_{ij} = d_{ji}$

Need not be a metric since

- it need not satisfy triangle inequality.

Minimax distance

Definition

$$d_{ij}^{MM} = \min_{r \in R_{ij}(G)} \Big(\max_{1 \leq l \leq |r|-1} d_{r(l)r(l+1)} \Big)$$

Minimax distance

Definition

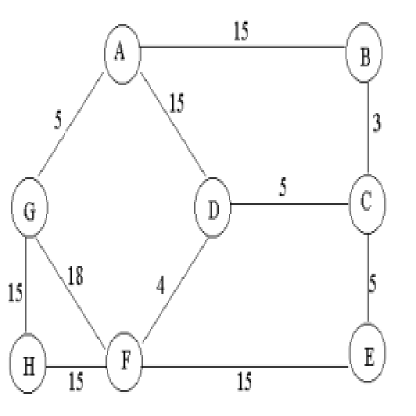
$$d_{ij}^{MM} = \min_{r \in R_{ij}(G)} \left(\max_{1 \le l \le |r|-1} d_{r(l)r(l+1)} \right)$$

where

- $R_{ij}(G)$- set of all possible paths between i and j,
- r- sequence of object indices,
- $r(l)$- l^{th} object in the path.

Minimax distance

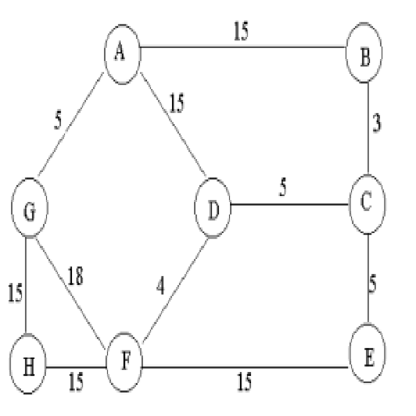
Example



- Base distance $d_{FG} = 18$.

Minimax distance

Example



- Base distance $d_{FG} = 18$.
- Minimax distance $d_{FG}^{MM} = 15$.

Minimax distance

Advantages

- Enable to compute the classes in a non-parametric way.

Minimax distance

Advantages

- Enable to compute the classes in a non-parametric way.
- Extract the class specific structures.

Minimax distance

Advantages

- Enable to compute the classes in a non-parametric way.
- Extract the class specific structures.
- Take into account the transitive relations.

Minimax distance

Advantages

- Enable to compute the classes in a non-parametric way.
- Extract the class specific structures.
- Take into account the transitive relations.
- Makes use of the distance function which is not necessarily a metric.

Classification using Minimax distance

Goal:

Classification using Minimax distance

Goal:

- Develop a general-purpose framework,

Classification using Minimax distance

Goal:

- Develop a general-purpose framework,
- To employ Minimax distances with any classification method,

Classification using Minimax distance

Goal:

- Develop a general-purpose framework,
- To employ Minimax distances with any classification method,
- That performs on numerical data.

Classification using Minimax distance

Plan of Action

1. Compute the pairwise minimax distance between all the objects i and j.

Classification using Minimax distance

Plan of Action

1. Compute the pairwise minimax distance between all the objects i and j.
2. Compute an embedding of these points in a new V.S.

Classification using Minimax distance

Plan of Action

1. Compute the pairwise minimax distance between all the objects i and j.
2. Compute an embedding of these points in a new V.S.
    - the pairwise squared Euclidean distance in the new space and
    - the pairwise minimax distance in the original space are equal.

Step 1- Computing pairwise minimax distance

Procedure:

- Build a minimum spanning tree (MST) over the graph,
- Compute the minimax distances over the MST.

Step 1- Computing pairwise minimax distance

Procedure:

- Build a minimum spanning tree (MST) over the graph,
- Compute the minimax distances over the MST.

Are the minimax distances over a graph and its MST same?

Equivalence of Minimax distance over a graph and its MST

Theorem

- Given a graph $G(O, D)$,

Equivalence of Minimax distance over a graph and its MST

Theorem

- Given a graph $G(O, D)$,
- for every pair of objects i, j belongs to O,

Equivalence of Minimax distance over a graph and its MST

Theorem

- Given a graph $G(O, D)$,
- for every pair of objects i, j belongs to O,
- their Minimax distance over G and MST constructed on that is identical.

Equivalence of Minimax distance over a graph and its MST

Theorem

- Given a graph $G(O, D)$,
- for every pair of objects i, j belongs to O,
- their Minimax distance over G and MST constructed on that is identical.

Outcome

Equivalence of Minimax distance over a graph and its MST

Theorem

- Given a graph $G(O, D)$,
- for every pair of objects i, j belongs to O,
- their Minimax distance over G and MST constructed on that is identical.

Outcome

- The Minimax distance for the Minimum spanning tree is

$$d_{ij}^{MM} = \max_{1 \le l \le |r_{ij}|-1} d_{r_{ij}(l) r_{ij}(l+1)} \ .$$

Equivalence of Minimax distance over a graph and its MST

Theorem

- Given a graph $G(O, D)$,
- for every pair of objects i, j belongs to O,
- their Minimax distance over G and MST constructed on that is identical.

Outcome

- The Minimax distance for the Minimum spanning tree is

$$d_{ij}^{MM} = \max_{1 \leq l \leq |r_{ij}|-1} d_{r_{ij}(l) r_{ij}(l+1)} \ .$$

- Equivalent of getting the maximum edge weight in the path between them.

Step 2- Embedding of pairwise Minimax distances

Given

D^{MM}- pairwise minimax distance between the objects.

Step 2- Embedding of pairwise Minimax distances

Given

D^{MM}- pairwise minimax distance between the objects.

Aim

- To find an embedding of the objects into a vector space.

Step 2- Embedding of pairwise Minimax distances

Given

D^{MM}- pairwise minimax distance between the objects.

Aim

- To find an embedding of the objects into a vector space.
  - The pairwise minimax distance in the original space and pairwise squared Euclidean distance in the new space are same.

Does an L_2^2 embedding exist or not?

Step 2- Embedding of pairwise Minimax distances

Given

D^{MM}- pairwise minimax distance between the objects.

Aim

- To find an embedding of the objects into a vector space.
  - The pairwise minimax distance in the original space and pairwise squared Euclidean distance in the new space are same.

Does an L_2^2 embedding exist or not?

Theorem

- Given the pairwise distances D^{MM},

Step 2- Embedding of pairwise Minimax distances

Given

D^{MM}- pairwise minimax distance between the objects.

Aim

- To find an embedding of the objects into a vector space.
  - The pairwise minimax distance in the original space and pairwise squared Euclidean distance in the new space are same.

Does an L_2^2 embedding exist or not?

Theorem

- Given the pairwise distances D^{MM},
- the matrix of Minimax distances D^{MM} induces an L_2^2 embedding.

Computation of squared Euclidean embedding

Computation of squared Euclidean embedding

Step 1: Centering the distance matrix D^{MM}.

$$W^{MM} = -\frac{1}{2} A D^{MM} A \tag{1}$$

where $A = I_N - \frac{1}{N} e_N e_N^T$.

Computation of squared Euclidean embedding

Step 1: Centering the distance matrix D^{MM}.

$$W^{MM} = -\frac{1}{2} A D^{MM} A \tag{1}$$

where $A = I_N - \frac{1}{N} e_N e_N^T$.

Step 2: Singular Value Decomposition of W^{MM}

$$W^{MM} = V \Lambda V' \; . \tag{2}$$

Computation of squared Euclidean embedding

Step 1: Centering the distance matrix D^{MM}.

$$W^{MM} = -\frac{1}{2} A D^{MM} A \tag{1}$$

where $A = I_N - \frac{1}{N} e_N e_N^T$.

Step 2: Singular Value Decomposition of W^{MM}

$$W^{MM} = V \Lambda V^{'} \ . \tag{2}$$

Step 3: Calculating Minimax vectors

$$Y_d^{MM} = V_d (\Lambda_d)^{\frac{1}{2}} \ . \tag{3}$$

Experiments

Classification Methods

1. Logistic Regression
2. SVM Linear Kernel
3. Nearest Neighbour

Experiments

Classification Methods

1. Logistic Regression
2. SVM Linear Kernel
3. Nearest Neighbour

datasets

- Synthetic:
  1. Circular
  2. Annular
  3. Chess Board
- Real:
  1. Iris

Experiments

Base distances

1. Euclidean distance
2. $d(x, y) = \left| \, \|x\|_2 - \|y\|_2 \, \right|$
3. Manhatten distance

Experiments

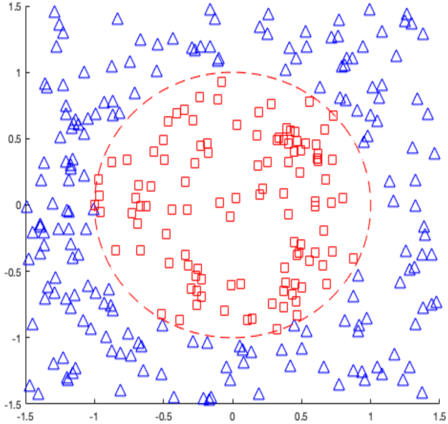
Base distances

1. Euclidean distance
2. $d(x, y) = | \, \|x\|_2 - \|y\|_2 \, |$
3. Manhatten distance

Performance parameters

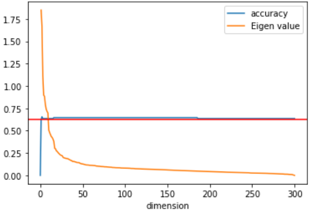
1. Accuracy of classification in the original space.
2. Accuracy of classification after embedding in the new space.

Circular Dataset

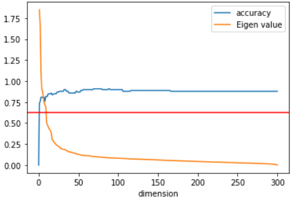


Circular Dataset

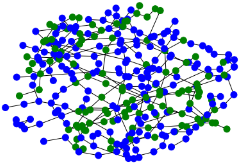
1. Euclidean Distance



(i) Logistic Regression



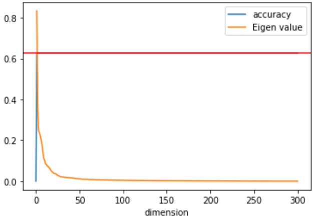
(ii) SVM Linear



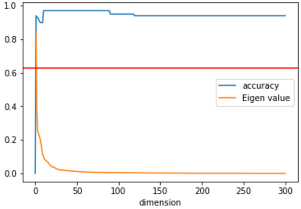
(iii) MST

Circular Dataset

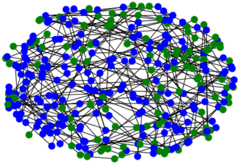
2. $d(x, y) = \left| \, \|x\|_2 - \|y\|_2 \, \right|$



(iv) Logistic Regression

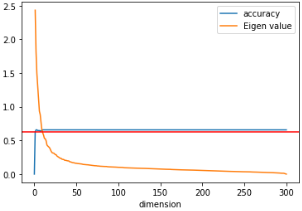


(v) SVM Linear

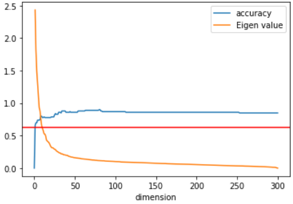


(vi) MST

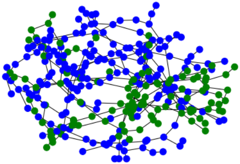
3. Manhatten Distance



(vii) Logistic Regression

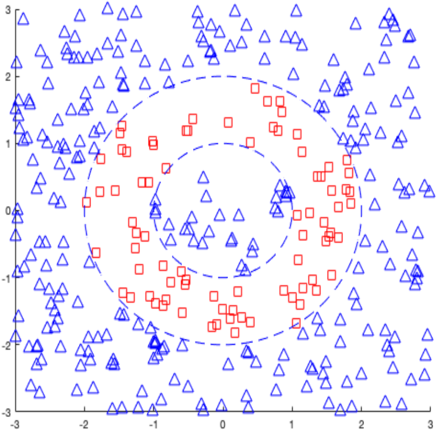


(viii) SVM Linear



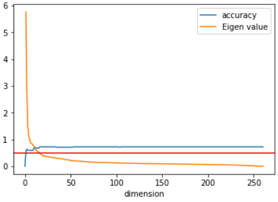
(ix) MST

Annular Dataset

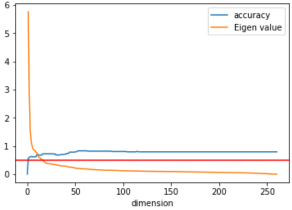


Annular Dataset

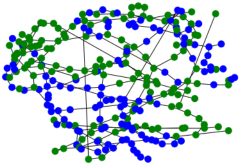
1. Euclidean Distance



(x) Logistic Regression



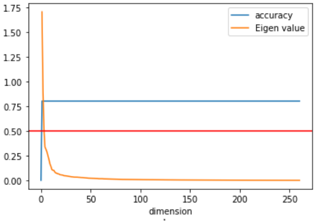
(xi) SVM Linear



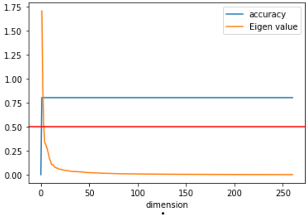
(xii) MST

Annular Dataset

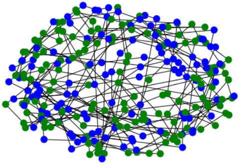
2. $d(x, y) = | \, \|x\|_2 - \|y\|_2 \, |$



(xiii) Logistic Regression

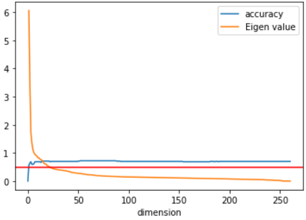


(xiv) SVM Linear

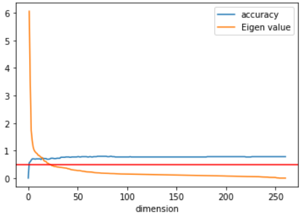


(xv) MST

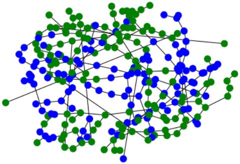
3. Manhatten Distance



(xvi) Logistic Regression



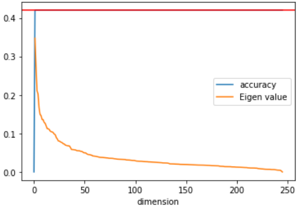
(xvii) SVM Linear



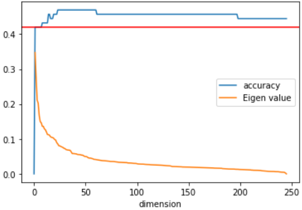
(xviii) MST

Chessboard Dataset

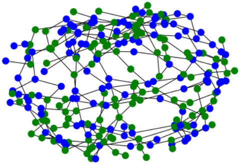
1. Euclidean Distance



(xix) Logistic Regression

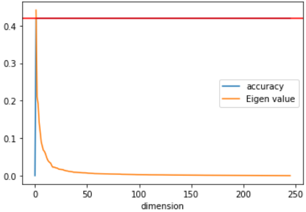


(xx) SVM Linear

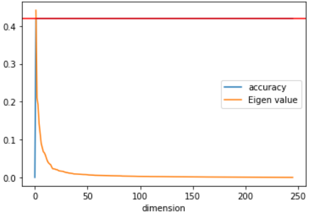


(xxi) MST

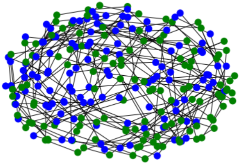
2. $d(x, y) = |\, \|x\|_2 - \|y\|_2 \,|$



(xxii) Logistic Regression

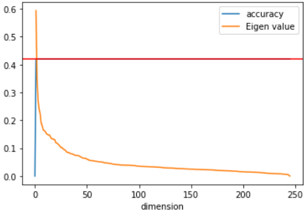


(xxiii) SVM Linear

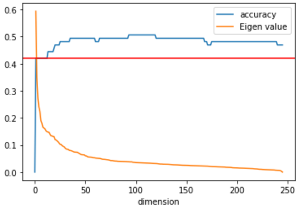


(xxiv) MST

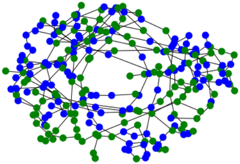
3. Manhatten Distance



(xxv) Logistic Regression

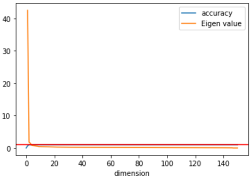


(xxvi) SVM Linear

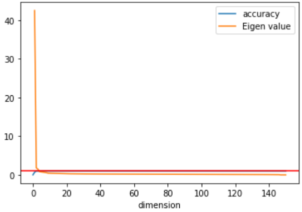


(xxvii) MST

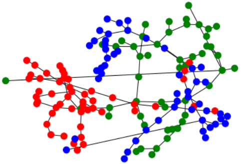
1. Euclidean Distance



(xxviii) Logistic Regression



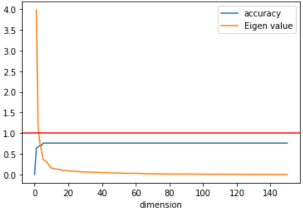
(xxix) SVM Linear



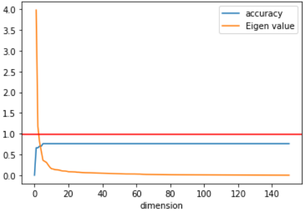
(xxx) MST

Iris dataset

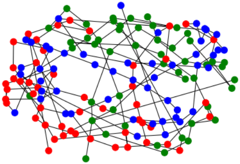
2. $d(x, y) = |\, \|x\|_2 + \|y\|_2 \,|$



(xxxi) Logistic Regression

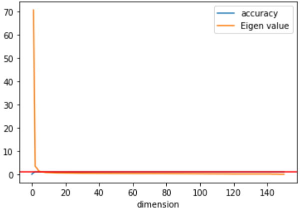


(xxxii) SVM Linear

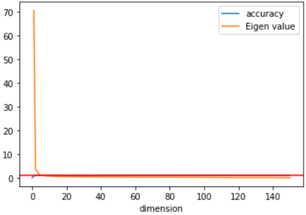


(xxxiii) MST

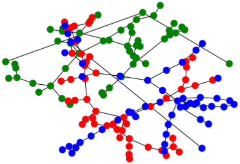
3. Manhatten Distance



(xxxiv) Logistic Regression



(xxxv) SVM Linear



(xxxvi) MST

Accuracy

Sno.	DataSet Name	distance function	Accuracy without embedding (Logistic regression)	Accuracy with embedding (logistic regression)	Accuracy without embedding (svm Lin)	Accuracy with embedding (svm lin)	Accuracy without embedding (NN)	Accuracy with embedding (NN)
1	Circular	Euclidean	0.626262626	0.646464646	0.626262626	0.888888889	0.95	0.923333333
		Norm 2	0.626262626	0.626262626	0.626262626	0.939393939	0.993333333	0.986666667
		Manhatten	0.626262626	0.656565657	0.626262626	0.858585859	0.96	0.933333333
2	Annular	Euclidean	0.5	0.709302326	0.5	0.790697674	0.896153846	0.884615385
		Norm 2	0.5	0.802325581	0.5	0.802325581	1	0.988461538
		Manhatten	0.5	0.697674419	0.5	0.76744186	0.888461538	0.884615385
3	Chess Board	Euclidean	0.419753086	0.419753086	0.419753086	0.456790123	0.771428571	0.734693878
		Norm 2	0.419753086	0.419753086	0.419753086	0.419753086	0.555102041	0.542857143
		Manhatten	0.419753086	0.419753086	0.419753086	0.49382716	0.783673469	0.742857143
4	Iris	Euclidean	1	0.96	0.98	0.98	0.96	0.926666667
		Norm 2	1	0.76	0.98	0.76	0.833333333	0.82
		Manhatten	1	1	0.98	1	0.953333333	0.94

Conclusion

Observations

- While LR and SVM largely benefit from the embedding based on Minimax distance,
  - NN does not.
  - Does not bring in more insight than the original pairwise distance.
- The embedding does not make use of the class information.

Conjecture

- A distance that understands the underlying relationship,
  - appropriate for embedding in an NN setting.
- Amount of enhancement obtained from embedding
  - dependent on the original base distance.

Thanks for your patient listening!!!