Possibilistic Fuzzy Neural Gas

Release 1.0.0

Amit Kumar

Jul 13, 2021

CONTENTS:

1 ALGORITHM DESCRIPTION **1**
- 1.1 Introduction . . . 1
- 1.2 Pseudo Code . . . 2

2 INSTALLATION REQUIREMENTS **3**
- 2.1 Libraries . . . 3
- 2.2 Execution . . . 3

3 Vectorial Data **5**
- 3.1 Example: Iris Data . . . 8

4 Relational **9**
- 4.1 Example: Random sample from three Gaussians . . . 12

5 Median **13**
- 5.1 Example: Random sample from three Gaussians . . . 16

Index **17**

CHAPTER

ONE

ALGORITHM DESCRIPTION

1.1 Introduction

Neural Gas is inspired by the Self-Organizing Map for finding optimal data representations based on feature vectors. The algorithm was named **Neural Gas** because of the dynamics of the feature vectors during the adaptation process, which distribute themselves like a gas within the data space. Among the variety of methods that has been developed for clustering, Neural gas uses *prototype-based* approach. Prototype-based technique tries to define clusters using small set of prototypes. Where each prototype tries to represent the group of data points based on similarity measure to the prototype which can be influenced by size and shape of the parameter. Example of prototype based clustering algorithms are: c-Means or k-means (where, c or k is the number of prototypes), Self-Organizing Map, Neural Gas etc.

The principle version of all these clustering methods is that each data point is represented uniquely by exactly one prototype. This is also known as crisp clustering. In real world most of the time data is overlapping, so a clustering method which uses soft assignment of data point to the prototypes can be very helpful in understanding the structure of the data. This is also known as clustering. Example of fuzzy clustering algorithms are: Fuzzy c-means, Fuzzy Self Organizing Map, Fuzzy Neural Gas etc.

Possibilistic Fuzzy Neural Gas(PFNG) helps in clustering when the data is overlapping. There are two kind of a assignments that are used:

`Probabilistic` – which ranges [0, 1] and for each data point to all the prototypes the values should sums up to 1.

`Possibilistic` – the value of the assignment decreases with the increase in the distance to the prototype.

PFNG can be used to cluster different kind of data:

`Vectorial data:` – Contains N x M-dimensional real valued vectors, and the distance measures used here is squared Euclidean distance.

`Relational data:` – It is a non vectorial data, the dissimilarity matrix **D** is provided which is of N x N dimension and the matrix is symmetric and complete. e.g. music, text, gene sequence etc.

`Median data:` –: Contains N x N dimensional data, the data has same properties as Relational data.

1.2 Pseudo Code

1. Set the number of prototypes N_p.
2. Initialise prototypes:
   - Vectorial: generate random vectors in data space.
   - Relational: intitalise random coefficients α_j.
   - Median: select random data samples from the data itself.

   repeat:

3 : calculate fuzzy assignment u_{ij}.

4 : calculate typicality t_{ij}.

5 : update probability using u_{ij} and t_{ij} from step 3 and 4
   - Vectorial: calculate w_j.
   - Relational: update coefficients α_j.
   - Median: select random data samples from the data itself.

   until: Convergence or number of epochs or manual stop.

CHAPTER

TWO

INSTALLATION REQUIREMENTS

2.1 Libraries

The basic requirements to run the code are:

- Minimum python 3.8 version.
- numpy 1.21.0
- matplotlib 3.4.2
- math
- imagio
- os
- shutil
- scipy.spatial.distance
- sklearn

2.2 Execution

- Copy the code.
- Go at the end of the code.
- Change the location of the data and change the number of prototypes.
- Run the code.

CHAPTER

THREE

VECTORIAL DATA

class Vectorial.**VectorialNeuralGas**(*data*, *epochs*, *generate_plot=False*, *col1=None*, *col2=None*)

Bases: object

Vectorial Neural Gas Clustering uses prototypes to represent the data by small number of vectors which are the points in data space. Each prototype tries to express the distribution of similar data points using the similarity measures. This algorithm uses membership and typicality values to find group with similar features when the data is overlapping. The prototype updates in this program is done explicitly.

updateMemTyp()

This function updates the membership values and typicality values after the every iteration of the prototype update.

class Vectorial.**Vectorial_Data**(*file*, *number_of_prototype*, *use_Columns*)

Bases: object

This class calculates various metric to perform task of Vectorial data clustering.

b

self.gamma is a constant.

beta_ij()

This function calculates the degree of belongingness of a data point to the prototypes.

var second_term: float value which contains value of typicality[i][l] to the power of typicality constant and then multiplied with scalar b.

var first term: float value which contains value of membership[i][l] to the power of membership constant and then multiplied with scalar a.

return: a N x K matrix.

create_animation(*name*)

This function creates an .gif file using the image saved at every epoch.

param name: name to be used to save the .gif file.

data

self.number_of_prototype int value which contains the number of protoype(say K) to represent the data.

distance_prototype(*j*, *l*)

This function calculates the rank of the lth prototype according to j^{th} prototypes.

param j: contains the index of j^{th} prototype. param l: contains the index of lth prototype.

var l: contains the 1 x M vector of lth prototype. var j: contains the 1 x M vector of j^{th} prototype. var rank_j: everytime adds 1, when the d(w_l,w_j) - d(w_l, w_k) > 0 else 0.

returns: the rank of lth prototype according to j^{th} prototype. rtype: int

gamma
self.typicality_matrix t_ij is N x K dimension matrix whose values are in range [0,1].

generate_plot(*col1*, *col2*, *filename*)
This function generates plot of the data as an image file.

param col1: first column to be used to plot the data. param col2: second column to be used to plot the data. param filename: file name to be used for the generated image.

var count: iterates through the length of the data. var n: contains the index of the nearest prototype.

generate_vectors()
This function iterates through number of prototypes K, to generate and appends the vector as an array in self.prototypes.

Returns a K x M dimension matrix

get_nearest_n(*point*)
This function iterates through data length and fetches the maximum value of a degree of affinity of a data to different clusters. param point: contains the index of the data.

var nearest: initialises nearest to the first value of Beta_ij as nearest. var nearest_prototype: contains the index of the prototype which is nearest.

return: prototype which is nearest to the data.

initializeMembershipMatrix()
Initializes the random values in range (0,1), with the condition that the sum of elements of any row is 1.

var random_num_list: contains the list of size K with random numbers in range [0.0 , 1.0). var summation: contains the sum of list. var temp_list: contains the list which sums up to 1.

return: a N x K membership matrix.

initializeTypicalityMatrix()
Initializes the random values in range [0,1), but the row wise element sum need not be equal to 1.

var random_number_list: contains the list of size K with random numbers in range [0 , 1).

return: a N x K typicality matrix.

length
self.mem is the hyper-parameter that controls how fuzzy the cluster will be, ranges usually between 1.2 to 2.0.

local_loss(*i*, *j*)
The function calculates the error sum, between data points at i^{th} index and prototypes at j^{th} index.

param i: contains the index of the i^{th} data. param j: contains the index of the j^{th} prototype.

var local_error: a double value which contains the sum of loss between data at i^{th} index and all the prototypes.

return: local loss between the data at index i and prototype at index j

mem
self.typ is the hyper-parameter that controls the degree of typicality values.

membership_matrix
self.generate_vectors generates random vectors.

neighborhood_function(*j*, *l*)
The function calculates range of prototype at j^{th} index in comparison with prototype at lth index using rank.

param j: contains the index of j^{th} prototype. param l: contains the index of lth prototype.

var rank_lj: int value returned by the ranking function. var sigma: a double value. var denominator: a double value. var numerator: int value containing (rank_lj)^2. var rank_neighborhood: double value containing the neighborhood value.

returns: the winner prototype which is needed to be updated. rtype: float

number_of_prototypes
self.length contains the length or the int value of the number rows(N) of the data.

print_dataset()

random_vectors()
This function generates random vector which has same number of column as of the data.

:returns a point of 1 x N dimension

scale()
This function preprocesses the data on one scale.

sig
self.a and self.b are positive independent scalars not necessary to sum up to 1.

sqEuclideanDistance(*i*, *l*)
The function calculates the squared Euclidean Distance between data at index i and prototype at index l.

param i: contains the index of the i^{th} data. param l: contains the index of the lth prototype.

var distance_il: calculates the squared euclidean distance between data at i^{th} and prototype at lth index.

return: distance between the i^{th} data and the lth prototype. rtype: float

typ
self.sig is a constant value used in calculation of neighborhood.

typicality_matrix
self.membership_matrix is N x K dimension matrix whose values are in range [0, 1] and for any i u_ij the values should sum up to 1.

updateMembershipMatrix(*i*, *j*)
This function updates the membership value of the data at index i according to the prototype at index j.

param i: index of the i^{th} data. param j: index of the j^{th} prototype.

var power_m: a float value. var i: contains data at i^{th} index. var j: contains prototype at j^{th} index. var numerator: calculates the numerator part of updating the membership equation. var denominator: calculates the denominator part of updating the membership equation. var div: contains the result after division of the numerator with denominator.

updatePrototype(*j*)
This function updates j^{th} the prototype explicitly as the distance measure chosen is squared Euclidean Distance.

param j: index of the prototype which has to be updated.

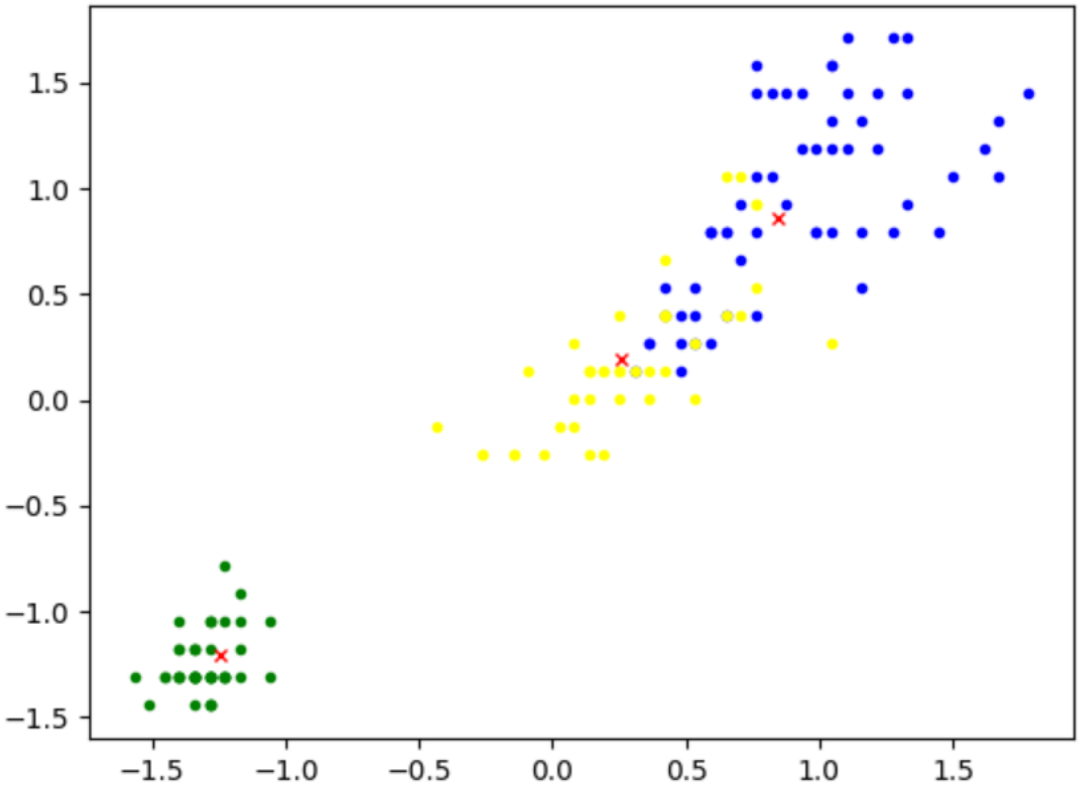
var num: calculates the numerator part of the prototype update equation. var den: calculates the denominator part of the prototype update equation.

updateTypicalityMatrix(*i*, *j*)
This function updates the typicality value of the data at index i according to the prototype at index j.

param i: index of the i^{th} data. param j: index of the j^{th} prototype.

var num: calculates the numerator part of updating the typicality equation. var den: calculates the denominator part of updating the typicality equation var div: contains the result after division of the numerator with denominator.

3.1 Example: Iris Data



CHAPTER

FOUR

RELATIONAL

class Relational.**RelationalNeuralGas**(*data*, *epochs*)

Bases: object

updateMembershipTypicality()

updateMemTyp(): updates the membership values and typicality values after the every iteration of the prototype update.

class Relational.**Relational_Data**(*file*, *number_of_prototype*)

Bases: object

Relational Neural Gas

Here, assumptions are made that there exists a non-linear mapping which projects the data into a Euclidean Space thus, there is a Dissimilarity matrix (interpreted Euclidean Space X_E), which contains the pairwise dissimilarities between the data objects. The matrix D is symmetric, where D_ij = D_ji and D_ii = 0. Also, unlike Vectorial data there is no direct prototype involved but, they are described indirectly by the coefficient Alpha_j so, by updating Alpha virtual prototypes are generated.

CalculateCentering(*data_r*, *data_s*)

Takes the indexes of the data and returns the value at that particular location.

Parameters

- **data_r** – index of Dissimilarity data.
- **data_s** – index of the Dissimilarity data.

Returns Value after double centering the data at index data_r and data_s.

DoubleCentering()

Double centering the data means any row or any column should sum to 0.

beta_ij()

@:var second_term: float value which contains value of typicality[i][j] to the power of typicality constant
and then multiplied with scalar b.

@:var first term: float value which contains value of membership[i][j] to the power of membership constant
and then multiplied with scalar a.

Returns a M x k matrix.

distanceDataNeuron(*i*, *j*)

The distances between the D_i and Alpha_j for Relational data can be calculated using this function.

Parameters

- **i** – index of the Dissimilarity matrix.
- **j** – index of the coefficient Alpha.

Returns Distance between the D_i and Alpha_j.

distance_prototype(*j*, *k*)
Calculates the distance between the Alphas.

Parameters

- **j** – j^{th} index of the Alpha.
- **k** – kth index of the Alpha.

Returns Distance between the Alpha at j and Alpha at k.

generate_coefficient()
This function generates the coefficient Alpha which should range (0,1], with the condition that the sum of every single row should be equal to 1.

Returns k x M matrix

generate_prototype()
The prototypes are updated indirectly using the updated Alpha values.

Returns The matrix of k x M of actual prototype.

initializeMembershipMatrix()
@:var random_num_list: contains the list of size K with random numbers in range [0.0 , 1.0). @:var summation: contains the sum of list. @:var temp_list: contains the list which sums up to 1.

Returns a N x K membership matrix.

initializeTypicalityMatrix()
@:var random_number_list: contains the list of size K with random numbers in range [0.0 , 1.0). :return: a N x K typicality matrix.

local_loss(*i*, *j*)

@:var second_term: float value which contains value of typicality[i][l] to the power of typicality constant and then multiplied with scalar b.

@:var first term: float value which contains value of membership[i][l] to the power of membership constant and then multiplied with scalar a.

Returns a M x k matrix.

neighborhood_function(*j*, *l*)

Parameters

- **j** – contains the index of j^{th} Alpha.
- **l** – contains the index of lth Alpha.

@:var rank_lj: int value returned by the ranking function. @:var sigma: a double value. @:var denominator: a double value. @:var numerator: int value containing (rank_lj)^2. @:var rank_neighborhood: double value containing the neighborhood value.

Returns the neighborhood range according to the winning Alpha.

normaliseAlpha()

Everytime the coefficients are updates, there is a need for normalisation of these coefficient so that the every row sum should be equal to one @:var sum_alpha: takes the sum of row

rank_ij(*j*, *l*)

calculates the rank of the lth Alpha according to j^{th} Alpha.

@:param j: contains the index of j^{th} Alpha. @:param l: contains the index of lth Alpha.

@:var prototype_l: contains the 1 x M vector with index j. @:var prototype_j: contains the 1 x M vector with index l. @:var rank: everytime adds 1, when the d(l, j) - d(l, k) > 0 else 0.

@:returns: the rank of lth Alpha according to j^{th} Alpha.

updateMembershipMatrix(*i*, *j*)

updates the membership of the data at index i according to the Alpha at index j.

Parameters

- **i** – index of the i^{th} data.
- **j** – index of the j^{th} Alpha.

@var power_m: float value. @var i: contains data at index i. @var j: contains neuron at index j. @var numerator: calculates the numerator part of updating the membership equation. @var denominator: calculates the denominator part of updating the membership equation. @:var div: contains the result after division of the numerator with denominator.

updateTypicalityMatrix(*i*, *j*)

Parameters

- **i** – index of the i^{th} data.
- **j** – index of the j^{th} neuron.

@var num: calculates the numerator part of updating the typicality equation. @var den: calculates the denominator part of updating the typicality equation @var div: contains the result after division of the numerator with denominator.

update_coefficient(*p*, *q*)

This function takes coefficient matrix and updates the values of the coefficient using the Stochastic gradient descent learning.

Parameters

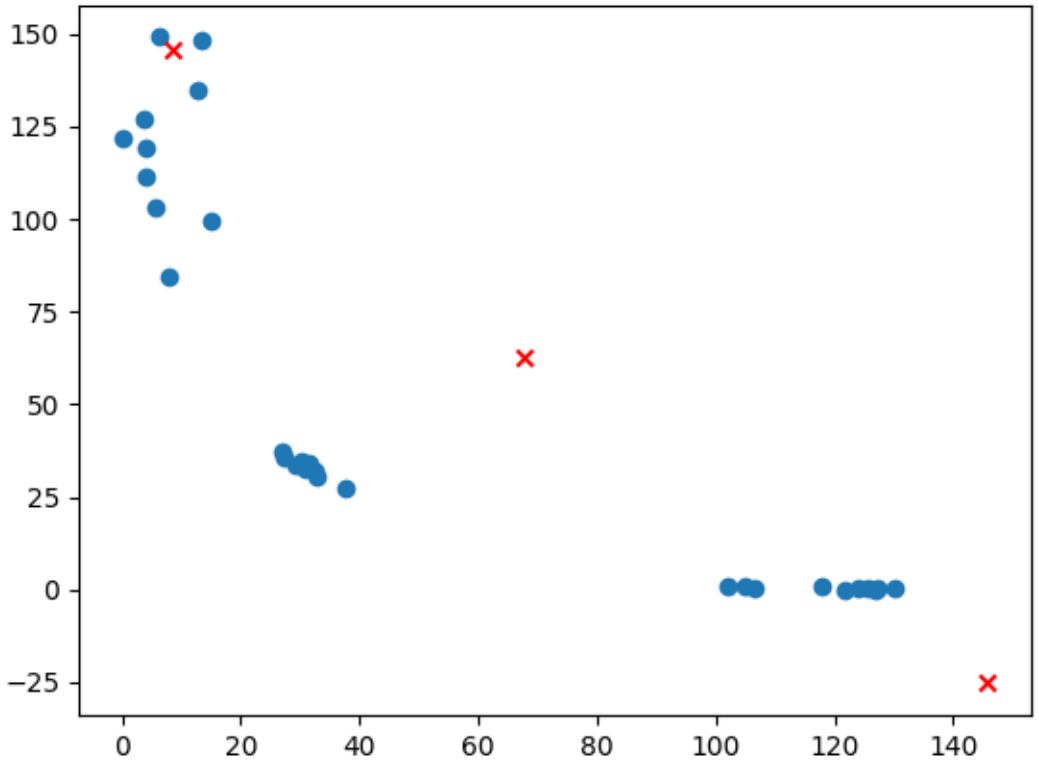
- **q** – index of the qth column of the coefficient matrix.
- **p** – index of the pth row of the coefficient matrix.

visualise_update()

When the coefficients gets updates, this functions visualises where the prototypes are moving.

Returns An image as output which contains scatter plots.

4.1 Example: Random sample from three Gaussians



CHAPTER

FIVE

MEDIAN

class Median.**MedianNeuralGas**(*data*, *epochs*)

Bases: object

Median Neural Gas

This is a reserved method in python classes. It is called as a constructor in object oriented terminology. This method is called when an object of Median_Data is created and it allows the class to initialize the attributes of the class.

para self: represents the instance of the class. param file: contains the link where, the data is stored. param number_of_prototype: predefined int type value which is required to represent the data.

updateMembershipTypicality()

This function updates the membership values and typicality values after the every iteration of the prototype update.

class Median.**Median_Data**(*file*, *number_of_prototype*)

Bases: object

This class calculates various metric to perform task of Vectorial data clustering.

CenteringData

"Vector of length N x 1 dimension.

D

Converts the data as an array

Distance

self.number_of_prototype int value which contains the number of protoype k to represent the data.

b

self.gamma is a constant.

beta

Initialisation of Membership and Typicality lists.

beta_ij()

This function calculates the degree of belongingness of a data point to the prototypes.

var second_term: float value which contains value of typicality[i][l] to the power of typicality constant
and then multiplied with scalar b.

var first term: float value which contains value of membership[i][l] to the power of membership constant
and then multiplied with scalar a.

return: a (N-k) x k matrix.

delDistance
self.mem is the hyper-parameter that controls how fuzzy the cluster will be, ranges usually between 1.2 to 2.0.

deleteDistance(*j*)
This function deletes the distances which are now acting as a prototype.

param j: Index of new prototype which is needed to be replaced from dissimilarity matrix.

gamma
initialises an (N-k) x k dimension matrix with zeros.

Type self.beta

generate_prototype()
This function chooses random k rows which act as an initial prototypes.

Returns The matrix of k x M of actual prototype.

initializeMembershipMatrix()
Initializes the random values in range (0,1), with the condition that the sum of elements of any row is 1.

var random_num_list: contains the list of size k with random numbers in range [0.0 , 1.0). var summation: contains the sum of list. var temp_list: contains the list which sums up to 1.

return: a (N-k) x k membership matrix.

initializeTypicalityMatrix()
Initializes the random values in range [0,1), but the row wise element sum need not be equal to 1.

var random_number_list: contains the list of size k with random numbers in range [0 , 1).

return: a (N-k) x k typicality matrix.

lenDist
self.length contains the the length of data excluding prototypes.

length
initialised self.Distance which acts as an array without prototypes.

local_loss(*i*, *j*)
The function calculates the error sum, between data points at i^{th} index and prototypes at j^{th} index.

param i: contains the index of the i^{th} data. param j: contains the index of the j^{th} prototype.

var local_error: a double value which contains the sum of loss between data at i^{th} index and all the prototypes.

return: local loss between the data at index i and prototype at index j

mem
self.typ is the hyper-parameter that controls the degree of typicality values.

membership_matrix
Initialises Centering N x N Matrix

neighborhood_function(*j*, *l*)
The function calculates range of prototype at j^{th} index in comparison with prototype at lth index using rank.

param j: contains the index of j^{th} prototype. param l: contains the index of lth prototype.

var rank_lj: int value returned by the ranking function. var sigma: a double value. var denominator: a double value. var numerator: int value containing (rank_lj)^2. var rank_neighborhood: double value containing the neighborhood value.

returns: the winner prototype which is needed to be updated. rtype: float

newPrototype(*j*, *l_dash*)

This function generates float values which is thrn used to decide the next prototype using argmin.

param j: Index of prototype to be updated. param l_dash: data length

return: returns M x 1 vector.

number_of_prototypes

self.lenDist contains the length of the Dissimilarity matrix.

rank_ij(*j*, *l*)

calculates the rank of the lth Alpha according to j^{th} Alpha.

param j: contains the index of j^{th} Alpha. param l: contains the index of lth Alpha.

var prototype_l: contains the 1 x M vector with index j. var prototype_j: contains the 1 x M vector with index l. var rank: everytime adds 1, when the d(l, j) - d(l, k) > 0 else 0.

returns: the rank of lth prototype according to j^{th} Alpha.

sig

self.a and self.b are positive independent scalars not necessary to sum up to 1.

sumBetaTyp

Initialises the membership, typicality matrices and generates prototypes

typ

self.sig is a constant value used in calculation of neighborhood.

updateMembershipMatrix(*i*, *j*)

This function updates the membership value of the data at index i according to the prototype at index j.

param i: index of the i^{th} data. param j: index of the j^{th} prototype.

var power_m: a float value. var i: contains data at i^{th} index. var j: contains prototype at j^{th} index. var numerator: calculates the numerator part of updating the membership equation. var denominator: calculates the denominator part of updating the membership equation. var div: contains the result after division of the numerator with denominator.

updateTypicalityMatrix(*i*, *j*)

This function updates the typicality value of the data at index i according to the prototype at index j.

param i: index of the i^{th} data. param j: index of the j^{th} prototype.

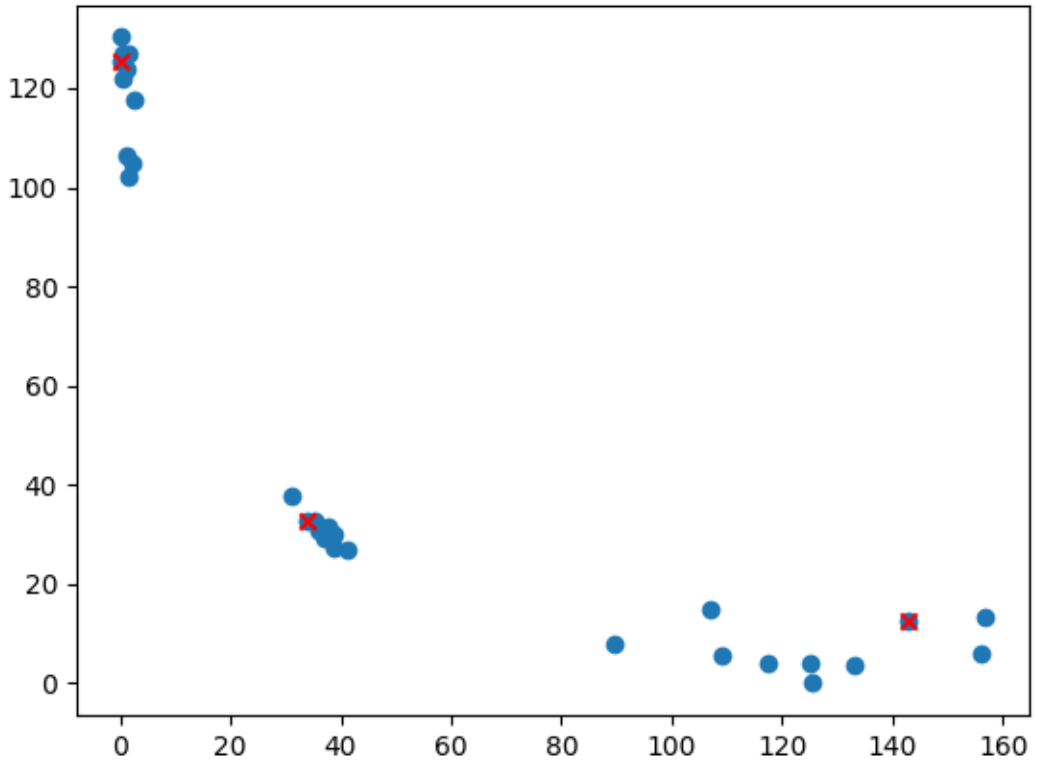
var num: calculates the numerator part of updating the typicality equation. var den: calculates the denominator part of updating the typicality equation var div: contains the result after division of the numerator with denominator.

visualise_update()

This function does the visualisation of movement of prototypes.

Returns An image as output which contains scatter plots.

5.1 Example: Random sample from three Gaussians



INDEX

B

b (*Median.Median_Data attribute*), 13
b (*Vectorial.Vectorial_Data attribute*), 5
beta (*Median.Median_Data attribute*), 13
beta_ij() (*Median.Median_Data method*), 13
beta_ij() (*Relational.Relational_Data method*), 9
beta_ij() (*Vectorial.Vectorial_Data method*), 5

C

CalculateCentering() (*Relational.Relational_Data method*), 9
CenteringData (*Median.Median_Data attribute*), 13
create_animation() (*Vectorial.Vectorial_Data method*), 5

D

D (*Median.Median_Data attribute*), 13
data (*Vectorial.Vectorial_Data attribute*), 5
delDistance (*Median.Median_Data attribute*), 13
deleteDistance() (*Median.Median_Data method*), 14
Distance (*Median.Median_Data attribute*), 13
distance_prototype() (*Relational.Relational_Data method*), 10
distance_prototype() (*Vectorial.Vectorial_Data method*), 5
distanceDataNeuron() (*Relational.Relational_Data method*), 9
DoubleCentering() (*Relational.Relational_Data method*), 9

G

gamma (*Median.Median_Data attribute*), 14
gamma (*Vectorial.Vectorial_Data attribute*), 5
generate_coefficient() (*Relational.Relational_Data method*), 10
generate_plot() (*Vectorial.Vectorial_Data method*), 6
generate_prototype() (*Median.Median_Data method*), 14
generate_prototype() (*Relational.Relational_Data method*), 10
generate_vectors() (*Vectorial.Vectorial_Data method*), 6
get_nearest_n() (*Vectorial.Vectorial_Data method*), 6

I

initializeMembershipMatrix() (*Median.Median_Data method*), 14
initializeMembershipMatrix() (*Relational.Relational_Data method*), 10
initializeMembershipMatrix() (*Vectorial.Vectorial_Data method*), 6
initializeTypicalityMatrix() (*Median.Median_Data method*), 14
initializeTypicalityMatrix() (*Relational.Relational_Data method*), 10
initializeTypicalityMatrix() (*Vectorial.Vectorial_Data method*), 6

L

lenDist (*Median.Median_Data attribute*), 14
length (*Median.Median_Data attribute*), 14
length (*Vectorial.Vectorial_Data attribute*), 6
local_loss() (*Median.Median_Data method*), 14
local_loss() (*Relational.Relational_Data method*), 10
local_loss() (*Vectorial.Vectorial_Data method*), 6

M

Median
    module, 13
Median_Data (*class in Median*), 13
MedianNeuralGas (*class in Median*), 13
mem (*Median.Median_Data attribute*), 14
mem (*Vectorial.Vectorial_Data attribute*), 6
membership_matrix (*Median.Median_Data attribute*), 14
membership_matrix (*Vectorial.Vectorial_Data attribute*), 6
module
    Median, 13
    Relational, 9
    Vectorial, 5

N

neighborhood_function() (*Median.Median_Data*

method), 14
neighborhood_function() (*Relational.Relational_Data method*), 10
neighborhood_function() (*Vectorial.Vectorial_Data method*), 6
newPrototype() (*Median.Median_Data method*), 15
normaliseAlpha() (*Relational.Relational_Data method*), 10
number_of_prototypes (*Median.Median_Data attribute*), 15
number_of_prototypes (*Vectorial.Vectorial_Data attribute*), 7

P

print_dataset() (*Vectorial.Vectorial_Data method*), 7

R

random_vectors() (*Vectorial.Vectorial_Data method*), 7
rank_ij() (*Median.Median_Data method*), 15
rank_ij() (*Relational.Relational_Data method*), 11
Relational
    module, 9
Relational_Data (*class in Relational*), 9
RelationalNeuralGas (*class in Relational*), 9

S

scale() (*Vectorial.Vectorial_Data method*), 7
sig (*Median.Median_Data attribute*), 15
sig (*Vectorial.Vectorial_Data attribute*), 7
sqEuclideanDistance() (*Vectorial.Vectorial_Data method*), 7
sumBetaTyp (*Median.Median_Data attribute*), 15

T

typ (*Median.Median_Data attribute*), 15
typ (*Vectorial.Vectorial_Data attribute*), 7
typicality_matrix (*Vectorial.Vectorial_Data attribute*), 7

U

update_coefficient() (*Relational.Relational_Data method*), 11
updateMembershipMatrix() (*Median.Median_Data method*), 15
updateMembershipMatrix() (*Relational.Relational_Data method*), 11
updateMembershipMatrix() (*Vectorial.Vectorial_Data method*), 7
updateMembershipTypicality() (*Median.MedianNeuralGas method*), 13
updateMembershipTypicality() (*Relational.RelationalNeuralGas method*), 9
updateMemTyp() (*Vectorial.VectorialNeuralGas method*), 5
updatePrototype() (*Vectorial.Vectorial_Data method*), 7
updateTypicalityMatrix() (*Median.Median_Data method*), 15
updateTypicalityMatrix() (*Relational.Relational_Data method*), 11
updateTypicalityMatrix() (*Vectorial.Vectorial_Data method*), 7

V

Vectorial
    module, 5
Vectorial_Data (*class in Vectorial*), 5
VectorialNeuralGas (*class in Vectorial*), 5
visualise_update() (*Median.Median_Data method*), 15
visualise_update() (*Relational.Relational_Data method*), 11