

## Algorithm: Multidimensional Scaling (MDS)

1.

Multidimensional Scaling (MDS) is a means of tool to visualize the level of similarity of individual cases of a dataset and it is able to translate the information about the pairwise distances among a set of objects or individual into a configuration of points mapped into an abstract Cartesian space according to the Wikipedia. However, the analytic tech which is a website specially about the machine learning algorithm, point out the purpose of Multidimensional to offer a vivid image of the pattern of proximities. In the other hand, the Chapter 435 of the official owner user manual states the MDS is a technique that will eventually build up a map to demonstrate the associated positions of the number of objects at the considerations of the table of the distances between them.

In the other hand, Joseph Imperial who is the author of the article *The Multidimensional Scaling (MDS) algorithm for dimensionality reduction* at the Medium website, Multidimensional scaling was called as Principal Coordinates Analysis (PCoA). This algorithm model specific stated mode will require to use between pairs of objects.

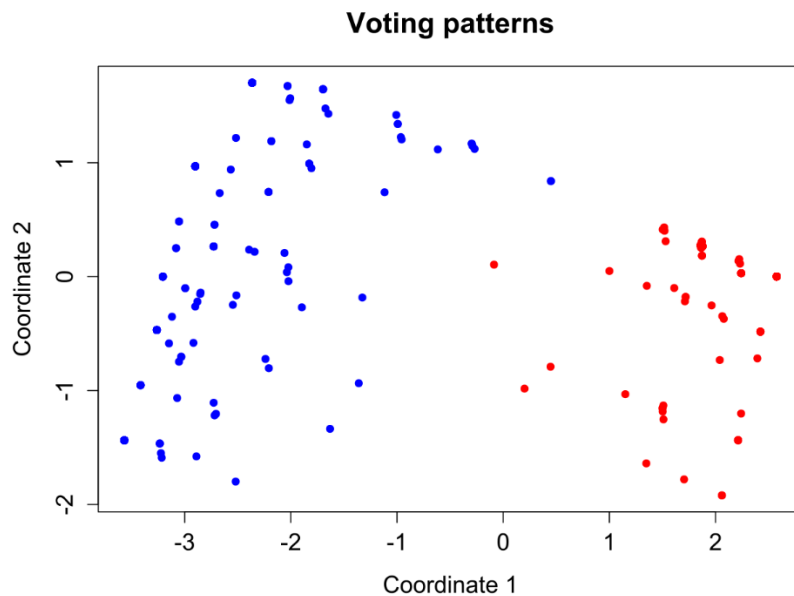
The core fundamental objective is these dissimilarities between points in these lower baskets of dimensional space.

2.

$$d_{rs}^2 = \sum_{i=1}^{n-1} \lambda_i (x_{ri} - x_{si})^2$$

Where,

$d_{i,j} :=$  distance between  $i$ -th and  $j$ -th objects.



3.

MDS analysis request the user to use the proximities. The proximities are another vision of the entire similarity between those elements inside the data. In this method, the MDS will try to find out the special spatial configuration of the elements because it needs the distances match exactly the same between the elements. In other words. The proximities should be as closely as possible. In addition, these proximities are going to be arranged in a square form matrix and be used primary between two different methods, direct or indirect methods.

Some real-world example for using MID analysis is such as the Ekman using the MIDS analysis machine learning algorithm to collect the data and find the appropriate perception of 14 different colors. Each single color was judged and determined by the respondent from having without similarities into the exact same one. Thus, if the color without any difference would be marked as 1, otherwise as 0.

In summary, MDS can be used in different types of data if it meets the requirements of the mix of the raw or transformed data will be appropriate for the multidimensional scaling method. It may not only limit to correlations, distances, proximities, similarities, multiple rating scales, preference matrices.

However, Multidimensional scaling methods may suit for the relational data matrices. In addition, it covers the symmetric, asymmetric, rectangular, square function matrices.

4.

The popular implication for Multidimensional scaling is particularly useful in the psychology industry. Because it can provide the sufficient information for the typical researchers such as the importance of the relationship. It will work every well within this area because the traditional psychosocial data probably is very large amounts and it touched most of different aspects.

However, the disadvantage of the Multidimensional scaling analysis method is also very distinct. It adds extra layer of subjectivity to the surface of psychological data. Usually, it requires users some kind of decision-making process. For example, the users need to determine which data appear on the table of the scale, which kind of multipliers should be applied to build up the rationale tables? However, it will add the risk of losing the risk since it may decrease the accuracy rate of the multidimensional scale.

MDS also faces the challenges such as incapable for analysis the large data sets. Second, the local minimum is another weakness for determine the best solution for those categorical data or nearby solutions.

5.

The following python library has been applied during the research:

- numpy

numpy python library can help users to access and manipulate the data values.

- Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

- Scikit-learn

Simple and effective tools for predictive data analysis and it can be reusable in various contexts.

6.

a. pyplot

matplotlib.pyplot is a collection of command style functions that make matplotlib work like MATLAB.

b.

Our dataset is the very basic dataset that was provided by Sk-learn python library. As the descriptions from the official website of Scikit-learn, the iris dataset is “classic and very easy multi-class classification dataset”. The dataset has 3 classes, 50 samples per class, 4 different dimensionalities, and real or positive features. The 3 classes are the three different types of irises, Setosa, Versicolour, and Virginica.

c.

The future research may also include the next step for the machine learning algorithm, such as expanding or deep research. The Nonmetric multidimensional scaling (MDS, also NMDS and NMS) is the method to differs in a various way to nearly the other ordination methods.

During the research path of the Jupyter notebook, our team found out we only performed the most basic application for the multidimensional scaling. The further implication includes the comparison with the principal component analysis, a large number of datasets with the numeric overwhelming features, and dataset with the natural clusters in multivariate space.

7.

Section	Links	Notes
1	<a href="https://en.wikipedia.org/wiki/Multidimensional_scaling">https://en.wikipedia.org/wiki/Multidimensional_scaling</a>	Wikipedia
1	<a href="http://www.analytictech.com/borgatti/mds.htm">http://www.analytictech.com/borgatti/mds.htm</a>	Tech Blog
1	<a href="https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Multidimensional_Scaling.pdf">https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Multidimensional_Scaling.pdf</a>	Software User manual
3	<a href="https://medium.com/datadriveninvestor/the-multidimensional-scaling-mds-algorithm-for-dimensionality-reduction-9211f7fa5345">https://medium.com/datadriveninvestor/the-multidimensional-scaling-mds-algorithm-for-dimensionality-reduction-9211f7fa5345</a>	Tech Blog
1	<a href="https://scikit-learn.org/stable/modules/generated/sklearn.manifold.MDS.html?highlight=mds#">https://scikit-learn.org/stable/modules/generated/sklearn.manifold.MDS.html?highlight=mds#</a>	
2	<a href="https://www.youtube.com/watch?v=Yt0o8ukIOKU">https://www.youtube.com/watch?v=Yt0o8ukIOKU</a>	
3	<a href="https://github.com/heucoder/dimensionality_reduction_alo_codes/blob/master/codes/MDS/MDS.py">https://github.com/heucoder/dimensionality_reduction_alo_codes/blob/master/codes/MDS/MDS.py</a>	
4	<a href="https://towardsdatascience.com/dimensionality-reduction-for-machine-learning-80a46c2ebb7e">https://towardsdatascience.com/dimensionality-reduction-for-machine-learning-80a46c2ebb7e</a>	
4	<a href="https://sciencing.com/advantages-disadvantages-multidimensional-scales-7320229.html">https://sciencing.com/advantages-disadvantages-multidimensional-scales-7320229.html</a>	Science website
5	<a href="https://www.geeksforgeeks.org/dimensionality-reduction/">https://www.geeksforgeeks.org/dimensionality-reduction/</a>	
6	<a href="https://pages.mtu.edu/~shanem/psy5220/daily/Day16/MDS.html">https://pages.mtu.edu/~shanem/psy5220/daily/Day16/MDS.html</a>	
7	<a href="https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py">https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py</a>	
7	<a href="https://strata.uga.edu/software/pdf/mdsTutorial.pdf">https://strata.uga.edu/software/pdf/mdsTutorial.pdf</a>	Nonmetric Multidimensional Scaling
8	<a href="https://github.com/swethapola/Multidimensional-Scaling/blob/master/131%20HW%204.ipynb">https://github.com/swethapola/Multidimensional-Scaling/blob/master/131%20HW%204.ipynb</a>	
9	<a href="https://www.programcreek.com/python/example/101758/sklearn.manifold.MDS">https://www.programcreek.com/python/example/101758/sklearn.manifold.MDS</a>	
10	<a href="https://www.displayr.com/what-is-multidimensional-scaling-mds/">https://www.displayr.com/what-is-multidimensional-scaling-mds/</a>	

## 8. Algorithm

In [7]:

```
# ADA_II
# HW4
#Qi Liu (Leader)
#Shiwen Chen
#Jiahua Chen
```

# Multidimensional Scaling

A powerful ordination method in inferential statistics / information visualization for exploring / visualizing the similarity (conversely the difference) between individual samples from a high dimensional dataset.

- beyond 2 or 3 features it is difficult to visualize the relationship between samples
- for 2 features we can easily visualize the relationships between samples with a scatter plot
- for 3 features we can either visualize in 3D or include color or matrix scatter plots

In [8]:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn import datasets,decomposition,manifold
```

## Loading iris dataset

In [9]:

```
def load_data():
    iris=datasets.load_iris()
    return iris.data,iris.target
```

## Adapting manifold.MDS function

In [10]:

```
def test_MDS(*data):
    X,y=data
    for n in [4,3,2,1]:
        mds=manifold.MDS(n_components=n)
        mds.fit(X,y)
        print('stress(n_components=%d):%s'%(n,str(mds.stress_)))
```

In [11]:

```
X,y=load_data()
test_MDS(X,y)
stress(n_components=4):15.509873133737447
stress(n_components=3):16.53620857516095
stress(n_components=2):156.8889190069542
stress(n_components=1):3323.02631881532
```

## Plotting the visual

In [12]:

```
def plot_MDS(*data):
    X,y=data
    mds=manifold.MDS(n_components=2)
    X_r=mds.fit_transform(X)

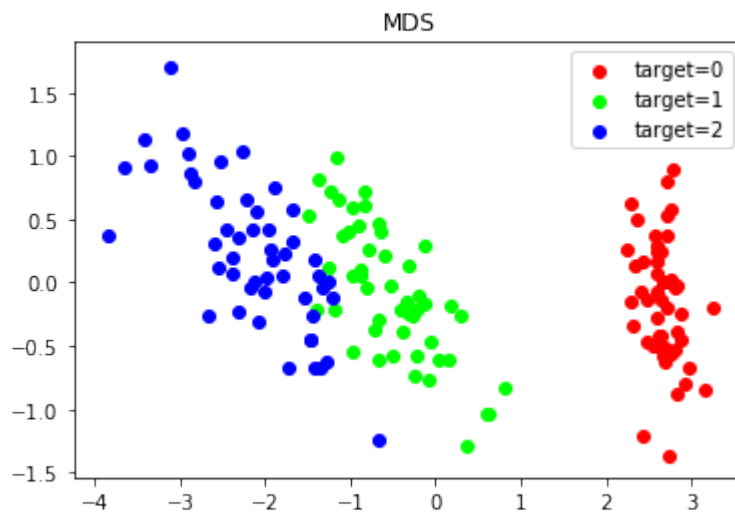
    fig=plt.figure()
    ax=fig.add_subplot(1,1,1)

    colors=((1,0,0),(0,1,0),(0,0,1),(0.5,0.5,0),(0,0.5,0.5),(0.5,0,0.5),
           (0.4,0.6,0),(0.6,0.4,0),(0,0.6,0.4),(0.5,0.3,0.2),)
    for label,color in zip(np.unique(y),colors):
        position=y==label

    ax.scatter(X_r[position,0],X_r[position,1],label='target=%d'%label,color=color)

    ax.set_xlabel('X[0]')
    ax.set_ylabel('X[1]')
    ax.legend(loc='best')
    ax.set_title('MDS')
    plt.show()

plot_MDS(X,y)
```



## Conclusion

This was a basic demonstration of multidimensional scaling. A lot more could be done:

- comparison to principal components analysis
- use of a dataset with larger number of features
- use of a dataset with natural clusters in multivariate space