International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Review Article)

Check for updates

Analysis of Manifold and its Application

Gyanvendra Pratap Singh and Shristi Srivastav *

Department of Mathematics and Statistics, DDU Gorakhpur University, Gorakhpur (UP) India.

International Journal of Science and Research Archive, 2024, 12(01), 394–404

Publication history: Received on 28 March 2024; revised on 04 May 2024; accepted on 07 May 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.12.1.0812

Abstract

Manifold learning is a field of study in machine learning and statistics that is closely associated with dimensionality reduction algorithmic techniques is gaining popularity these days. There are two types of manifold learning approaches: linear and nonlinear.

Principal component analysis (PCA) and multidimensional scaling (MDS) are two examples of linear techniques that have long been staples in the statistician's arsenal for evaluating multivariate data. Nonlinear manifold learning, which encompasses diffusion maps, Laplacian Eigenmaps, Hessian Eigenmaps, Isomap, and local linear embedding, has seen a surge in research effort recently. A few of these methods are nonlinear extensions of linear approaches. A nearest search, the definition of distances or affinities between points (a crucial component of these methods' effectiveness), and an Eigen problem for embedding high-dimensional points into a lower dimensional space make up the algorithmic process of the majority of these techniques. The strengths and weaknesses of the new method are briefly reviewed in this article. In the field of computer graphics, we utilize a particular manifold learning method was first presented in statistics and machine learning to create a global, Spectral-based shape descriptor.

Keywords: Manifold Learning; Isomaps; Embedding; Principal Component Analysis (PCA); Multi-dimensional scaling (MDS); Generative Topology Mapping (GTM)

1. Introduction

Differentiable manifold is an introductory course powerful framework manifold offer for dimension reduction. The key idea of dimension reduction finds to most succinct low dimension structure embedded to higher dimension structure.

A manifold is a higher dimensional extension of curves and surfaces. Manifold is a mathematical concept describing a space that locally resembles Euclidian space but may have a more complex global structure essentially. It can be used invarious fields. Global descriptors and manifold learning techniques aim to extract an informative and discriminative low dimensional vector of features by learning the geometry of a manifold.

*Corresponding author: Shristi Srivastav

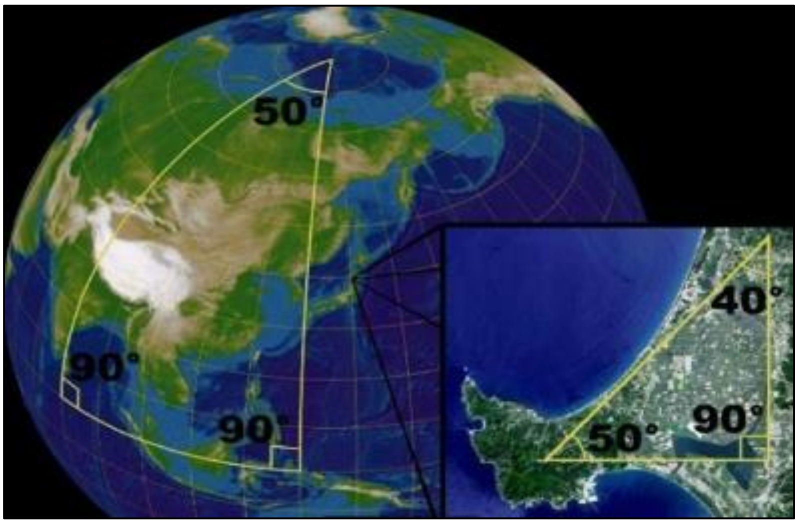

Figure 1 An example of Manifold

Example: Let us consider there is little black ant stroll along a manifold in 3D. This type of manifold could be twisty, curvaceous even have holes in it. Now here's rule; fromthe point of view of point of the black ant, all over the place it walks should see like a Flat Plane.

Does anyone recognize this rule? Since we all live in a multiverse, this app is probably the most relatable one you can find. One of the most basic examples of a 3D MANIFOLD is a sphere.

Manifold learning, on this other hand is set of technique used in machine learning to learn and represent the underline structure of data that may lie on a non-linear manifold it aims to capture the intrinsic dimensionality and relationships within the data is high dimensional potentially noisy manifold learning logrithams often transfer the data into a lower dimensional space were its structure more easily analized or visualized clustering classification and dimensionally reduction. In 2000 two cutting edge and original articles that appeared in the same scientific journal issue introduced the concept of multimodel learning. A class of techniques known as manifold learning or nonlinear dimensionality reduction aims to maintain the geometric and topological characteristics of a finite set using samples taken from a high-dimensional non-euclidean space.

The issue covered in these articles was how to retrieve data from a nonlinear low-dimensional manifold that is embedded in higher-dimensional ambient environment.

2. History of Manifold

The concept of a manifold has its roots in mathematics, particularly in geometry and topology. Manifolds were first formally introduced in the 19th century, with significant contributions from mathematicians like Bernhard Riemann and Henri Poincaré.

2.1. Early Concepts

The notion of a manifold began to take shape with the study of surfaces in three dimensional space. Mathematicians realized that surfaces such as spheres, cylinders, and tori could be described locally by simple Euclidean coordinates. This idea laid the foundation for the more abstract concept of a manifold.

2.2. Riemannian Manifolds

Bernhard Riemann made substantial contributions to the understanding of manifolds in the mid-19th century. He introduced the concept of a Riemannian manifold, which is a manifold equipped with a metric tensor. This allowed for the study of curved spaces and paved the way for Einstein's theory of general relativity.

2.3. Topology and Differentiable Manifolds

In the early 20th century, Henri Poincaré and others developed the field of topology, which deals with properties of spaces that are preserved under continuous deformations. This led to the concept of a topological manifold. Later, the notion of a differentiable manifold emerged, which is a manifold where smoothness can be defined.

Manifold learning is a relatively recent development in the field of machine learning and data analysis, emerging in the late 20th and early 21st centuries. Its key milestones.

2.4. Early Dimensionality Reduction Techniques

The need for dimensionality reduction techniques arose as datasets grew larger and more complex. Classical methods such as Principal Component Analysis (PCA) and Multi-Dimensional Scaling (MDS) were developed to reduce the dimensionality of data while preserving important properties. However, these methods often struggled with nonlinear and non-Euclidean data.

2.5. Nonlinear Dimensionality Reduction

In the late 1990s and early 2000s, researchers began exploring methods specifically designed to handle nonlinear data. One influential technique was Isomap (Isometric Mapping), introduced by Tenenbaum, de Silva, and Langford in 2000. Isomap aims to recover the underlying low-dimensional manifold structure of high-dimensional data by preserving geodesic distances embedded in the high-dimensional space.

2.6. Further Developments

Since the early 2000s, manifold learning has continued to evolve with the introduction of various algorithms and techniques, including t-distributed Stochastic Neighbour Embedding (t-SNE), Laplacian Eigenmaps, and Diffusion Maps. These methods offer different approaches to capturing the intrinsic structure of high-dimensional data and have found applications in fields such as computer vision, bioinformatics, and natural language processing. Over the years, mathematicians have extended the theory of manifolds to include more general structures such as complex manifolds, Symplectic manifolds, and algebraic varieties. These developments have found applications not only in mathematics but also in theoretical physics, computer graphics, and many other fields.

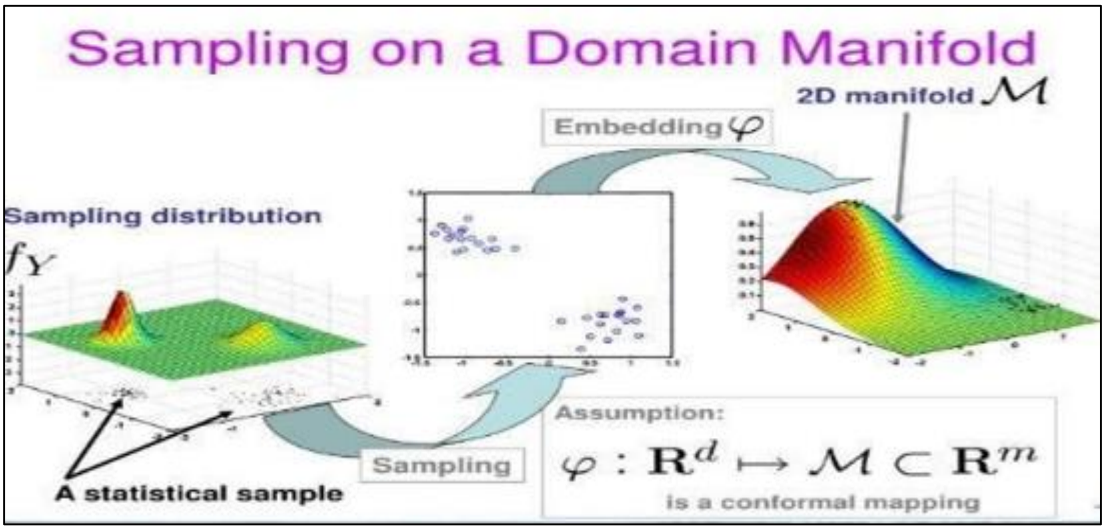

Figure 2 Sampling on a domain manifold

Overall, the historical background of manifold learning reflects a progression from linear techniques to more sophisticated methods capable of capturing the nonlinear and manifold structures present in complex datasets. Today, manifolds play a crucial role in various branches of mathematics and have applications in diverse areas of science and engineering. They provide a framework for understanding the geometry and topology of spaces of different dimensions and structures.

3. Preliminary

Differential Manifold, Kähler manifold, Topological manifold, C^K Manifold, Contact manifold, Riemannian manifold, Smooth manifold, Symplectic manifold, Finsler manifold, Algebraic manifold, Non-orientable manifold, Recurrent manifold, weakly symmetric manifold, Pseudo Symmetric manifold, Analytic Manifold and Complex Manifold.

3.1. Differential Manifold

A differential manifold is a mathematical object that locally resembles Euclidean space but may have a more complicated global structure. Formally, it is a topological space where every point has a neighbourhood homeomorphic to an open subset of Euclidean space R and these homeomorphic are compatible under change of coordinates. This structure allows for the study of smooth functions, vectors, and other geometric concepts.

In other word, Manifold is said to have n dimension if all of dimension n have connected components.Here also called 2 -dimensional manifold a surface, 1- dimensional manifold is a curve.

3.2. Embedding

An embedding is a way to represent data, typically high-dimensional data such as words, images or documents, in a lower-dimensional space where relationships between the data points are preserved. Some common types are showing below Isometric embedding, conformal embedding smooth embedding, topological embedding etc.

There are various types of embedding used in machine learning and natural language processing (NLP), including word embedding (such as word2vec, GloVe, and FastText), sentence embeddings (like universal sentence encoder and InterSent), and contextual embedding (such as BERT, GPT, and RoBERTa). Each type serves different purposes and has its own advantages and limitation.

Example: Subspaces of R to the power of n are among the most well-known examples of manifolds. Multivariable functions will be covered in a third semester applied calculus course in the United States. Vector fields, two dimensional surfaces, and volume under the following set of variables, for example, all of these wellbehaved objects are excellent examples of manifolds.

4. An Overview on Machine Learning:

When using Manifold learning approaches a learning process is usually involved with the goal of improving performance measure over time through experience in order to execute a task. Following the learning process the trained model can be used to classify predict or cluster new example based on the experience gained throughout the traning phase. Various statistics and mathematical models are used to calculate the performance of ML model and alogaritham. Fig (2) shows a typical ML approach.

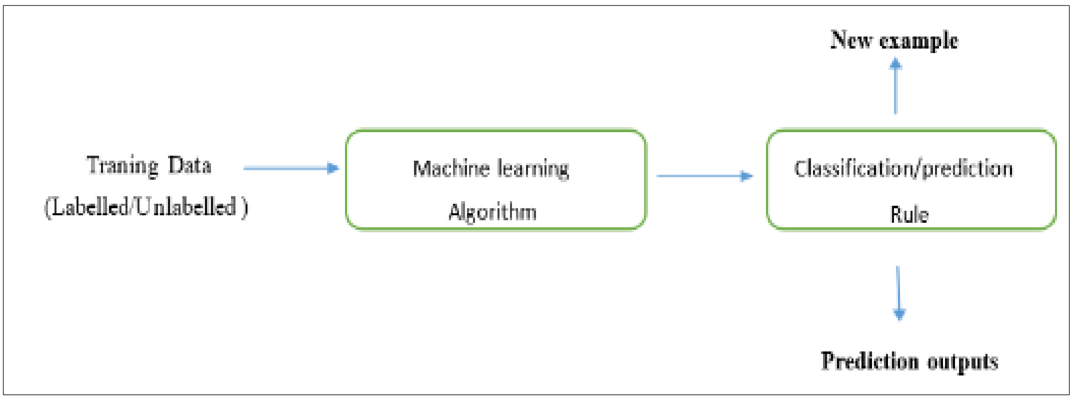

Figure 3 A typical machine learning Approach

4.1. Linear Manifold Learning

Linear manifold learning, also known as linear dimensionality reduction, refers to a class of techniques used in machine learning and data analysis to reduce the dimensionality of high-dimensional data while preserving as much relevant information as possible. The term "manifold" refers to a lower-dimensional subspace or surface embedded within the high-dimensional space where the data resides. Linear manifold learning methods seek to find a linear transformation of the original data into a lower-dimensional space, typically by identifying a set of linear combinations of the original features that capture the essential structure of the data. Examples of linear manifold learning techniques include Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). These methods are widely used for visualization, Noise reduction, and feature extraction in various applications such as image processing, Natural language processing, and signal processing.

4.1.1. Term 1: Principal Component Analysis (PCA):

Principal component analysis (PCA) is one of the most classical methods in dimensional reduction. Principal component analysis (PCA) is also called as the Karhunen-Loeve Transform (KLT) or singular value decomposition (SVD). The key idea of principal component analysis (PCA) is to find the low-dimensional linear subspace which Apprehend the greatest proportion of the modification within the data.

PCA considers the second order statistics of a random vector X belongs to $\Re^n$. Let $X_1,....X_N$ denote N samples from such a random vector. Let ohm denote the variance-covariance matrix of the random vector X, i.e., VAR(X) = E {[X-e(X)] [X-E(X)] ^{T}} =Ω.

Assume the symmetric and positive-semi definite matrix Ω has the following Eigen-decomposition:

$$\Omega=VDV^T,$$

Where V ϵ $\Re^{n..n}$is an orthogonal matrix ($V^TV = I_n$), and D is a diagonal matrix,

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$$

The diagonal entries of D, 0 ≤λn≤ λn-1≤…….≤λ1, are the ordered eigenvalues of Ω. The columns of V, V=(V1, V2…Vn)are the associated eigenvectors. From the following matrix computation, we observe thatλ_1,λ_2, …, and λ_k are the variances of the transformed random variables

$$V_1{}^TX,……V_2{}^TX……V_k{}^TX$$

$$\text{Cov}([V_1{}^T X, V_2{}^T,…….V_n{}^TX]) = \text{Cov}(V^TX)$$

$$= V^T \text{Cov}(X)\, V$$

$$= D$$

It is possible prove the projection X tensed to [V_1 ………..V_k]^T X from $\Re^n$ to $\Re^k$(k<n) keeps the greatest possible of variation in the data this sample only available the variance –covariance matix can be estimated

$$\sum_{i=1}^{N}(X_i - \bar{X})(X_i - \bar{X})^T$$

$$\text{Where } \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

PCA gives a natural dimension reduction. Consider an extreme case: if all the data lie in a low dimensional linear subspace of a very high dimensional space, then PCA will find such a linear subspace, because the variations in the directions that are orthogonal to the embedded linear subspace will be equal to zero.

The requirement that the embedded subspace be linear is a clear drawback of PCA; for a instance if the data are situated on a circle is 3D PC will not able to recognize such a structure. Mathematically speaking, PCA is a problem of finding the largest eigenvalues. We will demonstrate later that many algorithms ultimately lead to a matrix problem that is associated with eigenvalues, including MDS, LLE, Laplacian eigenmaps, and LTSA.

4.1.2. TERM: 2 Semi- Classical Method: Multi-Dimensional Scaling (MDS)

The acronym MDS refers to a collection of techniques with a broad variety of uses. The main idea is to map a high-dimensional space to a low-dimensional space in a way that optimally preserves the pairwise distances between the observed points. One logical illustration is to retrieve the relative locations of the cities based on the distances between them. Let's say that N cities' precise coordinates (locations) are lost. The driving distances between pairs of them are available to us, though. An array of these distances is created. With the help of this matrix, MDS is able to reconstruct a 2-D coordinate system with the positions of these cities. However, due to stiff motion a mix of rotation, shifting, and reflection.

4.2. Non linear manifold learning

Next, we go over a few algorithmic methods that have shown to be useful in the research of nonlinear manifold learning: Hessian Eigenmaps, Local Linear Embedding, Laplacian Eigenmaps, Diffusion Maps, Isomap, and the various nonlinear PCA variants.

4.2.1. TERM 1: Isomaps

All pairs of points between geodesic distances preserved are known isomaps. This the other method of non linear dimension reduction. The isometric feature mapping algorithm (Tenenbaum, de Silva, and Langford,

2000) Suppose that the smooth manifold M is a convex region of $\Re^t$ ($t \ll r$) and that the embedding ψ: M → Y is an isometry. This assumption has two key ingredients

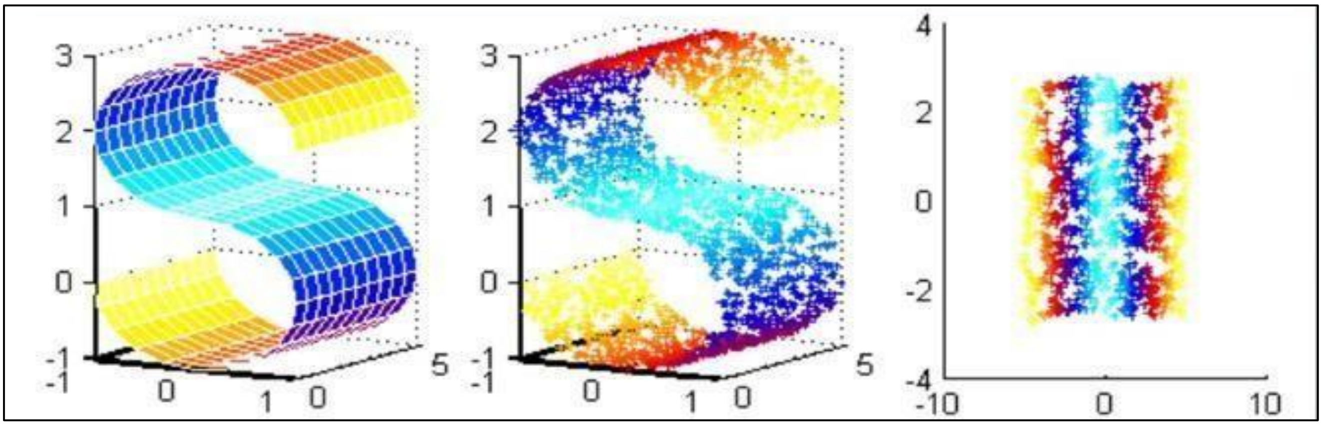

Figure 4 Left panel: The S-curve, a two-dimensional S-shaped manifold embedded in three-dimensional space. Right panel: 2,000 data points randomly generated to lie on the surface of the S-shaped manifold

- Isometry: Invariant under the geodesic distance the map ψ. For any pair of points on the manifold,z, z′ ∈ M, the geodesic distance between those points equals the Euclidean distance between their corresponding coordinates, x x′ ∈ X that is
  $d^m(z,z') = \|x-x'\|_X$
  where z = φ(x) and z ′ = φ (x ′).
- Convexity: The convex subset of a $\Re^t$ is the manifold M.

4.2.2. TERM2: Generative Topological Mapping (GTM)

An inspiring nonlinear dimension. Dimension reduction method All pairs of points between geodesic distances preserved are known Isomaps is known as GTM. While GTM formulation emphasis certain important elements of contemporary dimension reduction they do not contain the method that will be introduced later. Let t be the data space and X be a point in latent space. Let t_1 t_2…….t_n indicated that the observed point $t_i$is generated based on the following:

- First there is a quantity x_i associated with ti in the latent space has much lower dimension than the data space.
- There is a mapping x tending to y (x, U) that has complete column rank in its Jacobin and its continuously differentiable from the latent space to the data space. The parameters of this mapping are indicated by the letter

U. In reality, one can assume that the picture y (x, U) for any x form a low dimensional manifold in the data space.

- Consider that the observation t_i is generated according to the model

$$t_i = y(x_i ;U) + \varepsilon_i ,i=1,2,\ldots\ldots N$$

Thus, GTM takes an implicit manifold to exist. U and β are parameters that are unknown. Although they exist, the latent variables are not known x_i. The authors of GTM presented an EM based method to estimate the above model by assuming a particular distribution for the x_i's and putting the problem in a Bayesian model estimating framework (Bishop, Svensen, and Williams, 1998). The process of determining a maximum a posteriori (MAP) estimate results in the dimension reduction.

GTM considers a prior p(x) for the x_i 's. This prior is a sum of a finite number of Dirac functions, i.e.

$$p(x) = \sum_{i=1}^{k} \delta\,(x - \overline{x_i})$$

Where x_1, x_2, …, x_k are k given points in the latent space. According to the previous way of generatingt_i, there is a probability density function for t: p (t|x; U, β). The density function on the data space is simply

$$p\ (t/U, \beta) = \int p\ (t\ /\ x, U, \beta)\ p(x)dx$$

Given that p(x) is a sum of k Dirac functions, we have

$$p\ (t/U, \beta) = \sum_{i=1}^{k} p\ (t\ \overline{|}\overline{x_i}, U, \beta)$$

The principle of maximum likelihood estimation (MLE) is to find U and β such that the log-likelihood function

$$\sum_{j=1}^{N} \ln p\ (t_j \mid U, \beta)$$

is maximized. The authors of GTM (Bishop, Svensen, and Williams, 1998) proposed an expectation maximization (EM) approach to estimate U and β. Here we omit some of the technical details regarding how to choose the functional classes in the nonlinear mapping. The numerical solution of GTM is based on a strong assumption on the prior.

The way the EM algorithm is applied appears haphazard. Additionally, it is challenging to defend GTM's performance. In actuality, GTM can only be built as an alternative to self-organizing maps (SOM) in a few specific situations, such as clustering. The probabilistic approach, however, is consistent with other data analysis models.

4.2.3. TERM2: Locally Linear Embedding (LLE):

The LLE algorithm nonlinear dimensionality reduction is comparable to the isomap algorithm in several aspects, but we consider LLE to be an approach rather than the global approach represented by isomap since it aims to maintain local neighbourhood information on the manifold. Locally Linear Embedding (LLE): Another important milestone in manifold learning is the development of Locally Linear Embedding (LLE) by Roweis and Saul in 2000. LLE seeks to preserve the local relationships between data points, assuming that the data lie on or near a low-dimensional manifold embedded in the high-dimensional space.

5. Application of manifold

5.1. Mesh Generation: Modelling with multiple charts

It could be challenging to obtain enough coordinate precision far from the origin since the distances between floating point numbers grow with their magnitude. Avoiding circumstances where there are minute details that are distant (in a coordinate sense) from the source is an easy strategy to solve this issue. With a single coordinate system, this may be exceedingly challenging, but Generalization C states that the manifold directs us to divide the domain into subdomains, cover each subdomain with its create your own coordinate system, then position each coordinate system's origin to maximize precision. Multiple charts are helpful not only for mesh generation but also for enhancing accuracy other calculations made throughout solution process. A multiple chart example is presented in Figure This tactic is also discussed.

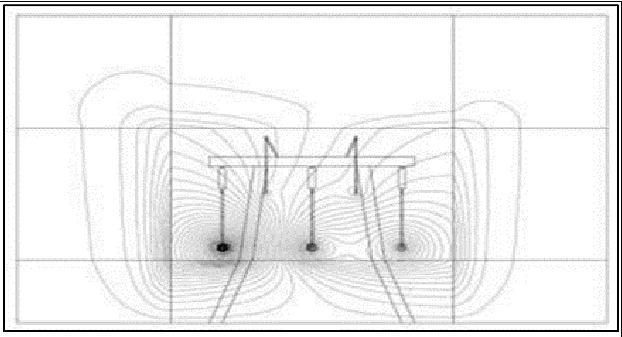	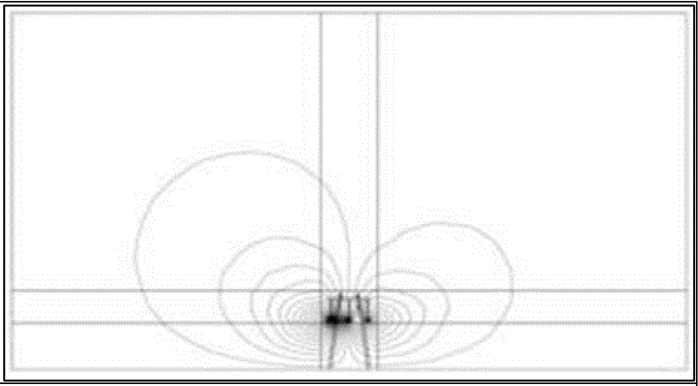
Figure 5 Isovalue lines of potential of the power in the non-standard parameterisation	**Figure 6** Isovalue lines of potential of the power line in the Line Standard parameterisation

This is the one and the same field as in Fig (4) but just shown with another chart.

Let's talk a little more about the multiple charts-strategies actual use. Although it will limit us to manifolds that can be embedded into the n-dimensional space, it is reasonable to assume that the user first provides, or at least implicitly assumes, a single Rn chart that is a standard parameterization and covers the entire manifold. The material parameters are given with respect to such chart.

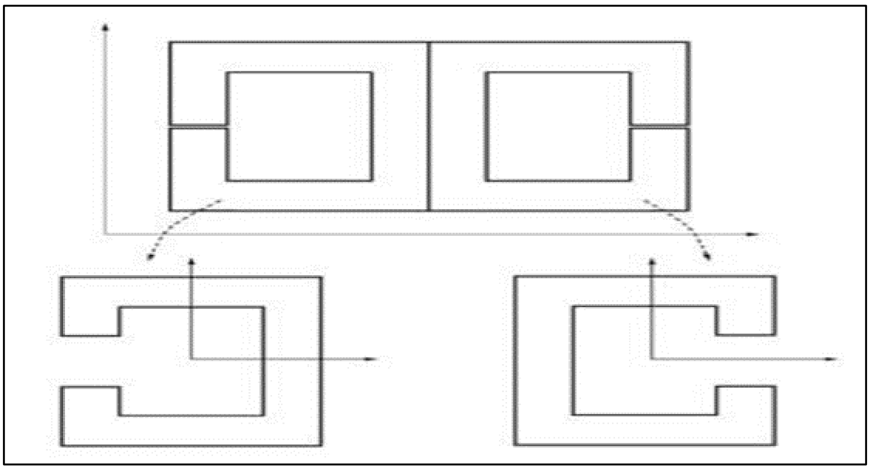

Figure 7 Example of multiple chart

Top: standard parameterization that covers whole manifold

Bottom: Two charts that span half of the domain are required to overlap only at their shared border according to convential parameterization, which also causes the origin and scale to be altered.

5.2. Machine learning in agriculture

A total of twenty-one articles were published in the journals Computer and Electronics in Agriculture,

Sustainability Real-Time Imagining, Precision Agriculture, Earth Observations.

Scientific Reports, Computers in Industry, and Sensors. Six articles were published in the journal Bio Systems Engineering. The majority of the papers are about using machine learning (ML) in crop management, although eight of the articles are about using ML in livestock management, four are about using ML in water management, and four are about using ML in soil management. shows how the articles are distributed in relation to the specified subcategories and these application domains.

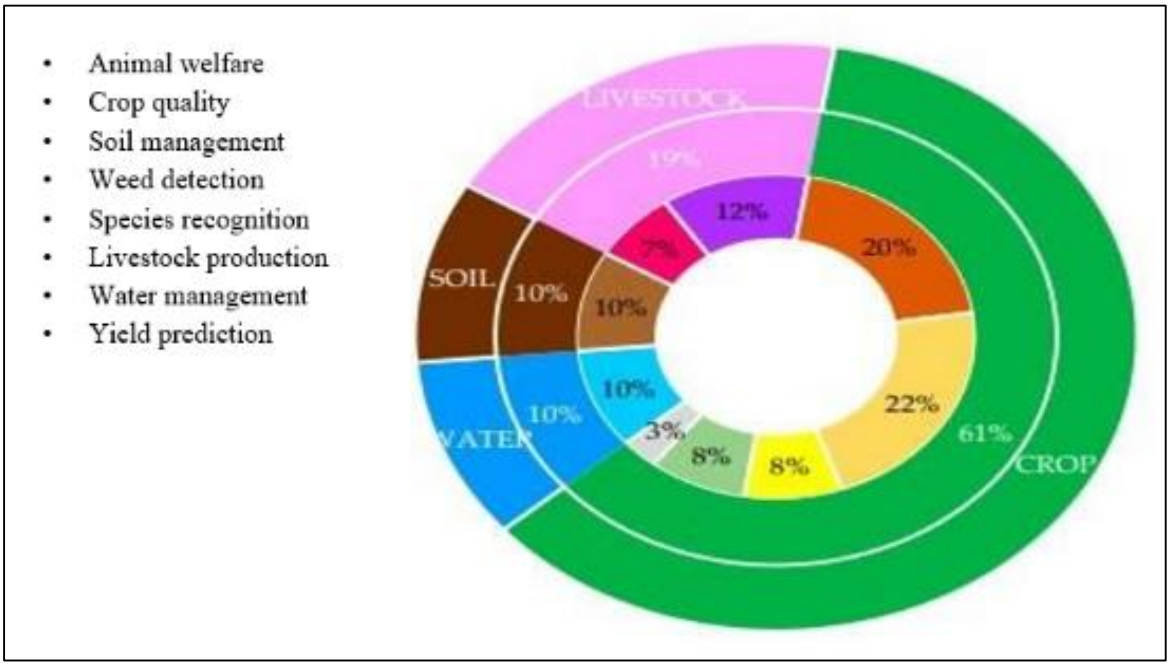

Figure 8 Pie chart presenting the papers according to the application domain

More precisely, five machine learning models were used in crop management techniques, with ANNs being the most widely used model (with wheat being the most common crop at hand). Four machine learning models were applied to the livestock management category; SVMs, which represent the most common type of livestock at hand cattle were the most widely used model. Two machine learning models were used for water management, specifically evapotranspiration estimation, with artificial neural networks (ANNs) being used most frequently.

5.3. Forecasting the course of brain tumors using a manifold learning algorithm

In 2011, the Texas University medical imaging department, in collaboration with the Anderson MD Oncology Center in Texas, unveiled a novel approach to multimodal learning. This study aims to propose a suitable technique for precisely diagnosing a low dimensional manifold associated with desired data structure.

The project's desired data collection is crucial for the diagnosis of brain tumors. As a result, MRI scan collections are included in data collecting. The goal of this study is to identify a manifold with a lower dimension that represents tumor, recovered, and healthy tissues. Furthermore, survey and research in the area of determining the interaction between tumor and healthy tissues are our top priorities. Through the charting of the manifold bridge that connected to Tumor progression can be seen in up to two consecutive MRI pictures, aids in overseeing the patient’s treatment regimen in this way. The theory put out in article is guaranteed and supported by the early stages of this study's results. In the lower space, manifolds associated with tumor and healthy tissues are detachable. Additionally, between these manifolds, the manifold associated with the progression of tumor tissue is located closer to the tumor tissue.

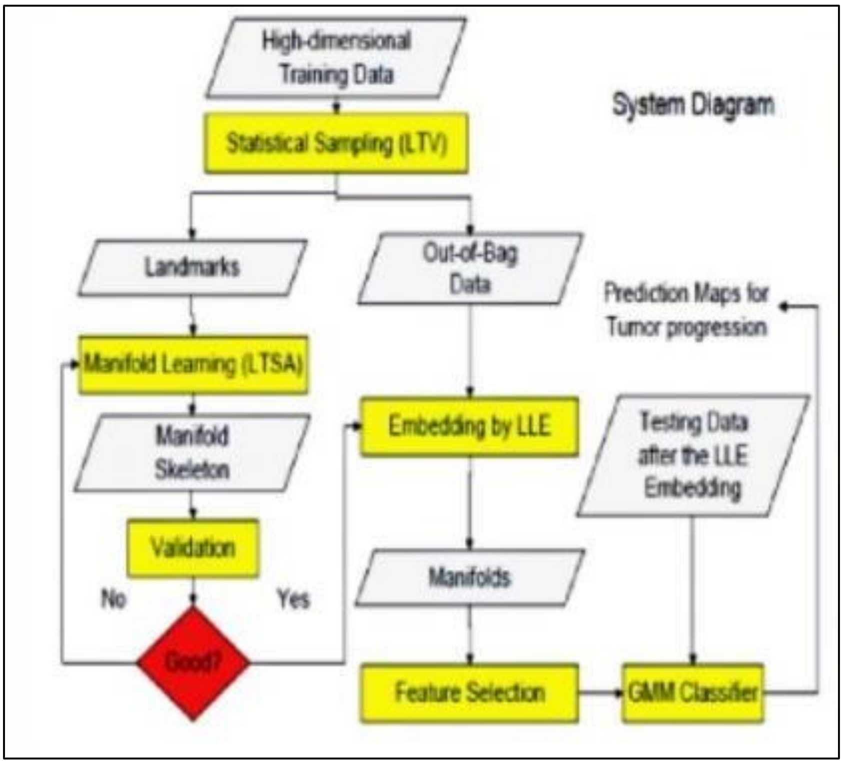

Figure 9 Block diagram of the application of Manifold learning for detection of Tumor progression

6. Conclusions

Learning the structure of a manifold is important when dealing with high dimensional data, like those from images and videos, that lie on or near a manifold of a lower dimensional space. This chapter provides an overview of the various methods that have been proposed for manifold learning. First we review the notations of a smooth manifold using fundamental concepts from topology and differential geometry to learn a linear manifold. Next we describe the global in imbeddings algorithm of principal component analysis and multidimensional scaling in the certain situations. Finally, we describe how linear methods will work to find the structure of a curved or non-linear manifold in certain situations.

We believe that by providing a thorough overview of manifold-based learning methods and highlighting their mathematical formulations, we will shed light on the ways in which these approaches are similar to an another and highlight the shared theoretical framework that will serve as the foundation for research in this field we also hope that this article will inspire new avenues for research in this field and draw attention to the theoretical analysis of manifold-based learning methods.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Alan Julian Izenman (2012).Introduction of manifold learning.

[2] Borg, I. and Groenen, P. (1997). Modern Multidimensional Scaling: Theory and application. Springer Verlag, New York.

[3] Costa J.A., Patwari, N And Hero, A.O.(2004).Distributed multidimensional scaling with adaptive weighting for node localization in sensor networks, Submitted to ACM trans. on, Sensor Networks, June.

[4] Coopersmith, E .J.; Minsker, B.S,; Wenzal ,C.E.; Gilmore , B.J.Machine Learning assessments of soil drying for agricultural planning . Comput. Electron .Agric (2014)104, 93-104. MDPI Journals, Sensor of volume 18.

[5] Dimensionality reduction Science(2000), 290:2319-2323.

[6] Institute for Laboratory Animal Research, National Research Council (2000). Humane endpoints for animals used in Biomedical Research and testing. ILAR journal, 41:59-123.

[7] International journals of Biomedical Engineering and Science (IJBES), Vol 1, NO. 2(July 2014).

[8] Kang.j Schwartz R; flickinger J; Beriwal machine learning approach for predicting radiation.

[9] Ma y Fu Y,eds. (2012). Manifold learning theory and its application.

[10] Ojo, R., Seriki, S., Wang, D. E.., Mhya, V. I. and Hyelni, T. (2015). Multifaceted Health benefits of Mangifera indica leaves (mango. Journal of pharamacy and Biological Science.5 (2):67-80.

[11] Protus, A. T., Gabriel, A.A.M, Tchangwue, D. A, Thokouaha, L.R. and Mengue, N. (2012). Growth yield and physiological of mango (Mangifera indica) cultivars as affected by polyembroymic rootstock. Journal of National product and plant Resources 5(5):617-627

[12] Rasouli, H., Fazaei, M. H. and Khodarahmi, R. (2017). Polyphenol and their benefits: A review of International Food Properties 20:38-42.

[13] Shah, K. A., Patel, M. B., Patel, R. J. and Parmar, P. K. (2010). Mangniferaindica (mango).Pharmacgen., 5(2):42-48.

[14] Singleton, p. (199). bacteria in biology, New York: John Wiley and Sons Ltd., P 337

[15] Tenenbaum JB, de silva V, Langford JC. A global geometric framework for nonlinear.

[16] The Institution of engineering and technology 2008 published in IET Science, Measurement and Technology Revised on 15th November 2007.

[17] Xiaoming Huo, Xuelei (sherry) Ni, and Andrew K. Smith (2007) Survey of manifold based learning Method.