أستخدام متعدد القياس الأعتيادي والشبكات العصبية لتقليل الأبعاد دراسة محاكاة

محمد أحمد جاسم
أسماء غالب جابر
جامعة بغداد/كلية الأدارة والأقتصاد/قسم الاحصاء

المستخلص:
يستعرض هذا البحث أسلوب قياس متعدد الأبعاد (Multidimensional Scaling - MDS) كأداة إحصائية لتحليل البيانات عالية الأبعاد وتمثيلها في فضائات ذات أبعاد أقل بهدف فهم العلاقات والأنماط فيها.يركز البحث بشكل خاص على طريقتين وهما قياس متعدد الأبعاد المتري:الحل الكلاسيكي((Metrice MDS والقياس متعدد الأبعاد بأستخدام الشبكات العصبية (NN-MDS) .
لتقيم أداء هاتين الطريقتين،تم اجراء دراسة محاكاة بأستخدام لغة برمجة الإحصائية MATLAB .تضمنت التجربة سيناريوهات متعددة شملت أحجام عينات مختلفة (50,100,250,500,1000) ،وعدد أبعاد أصلية (10,20,25,35,40)،مع أستهداف تقليص الى(1,2). واستخدم معيار الاجهاد (Stress Criterion) لتقيم مدى قدرة كل طريقة على الحفاظ على المسافات الأصلية بين النقاط بعد عملية التقليص.
أظهرت نتائج المحاكاة بشكل consistent أن طريقة القياس المتري(Metrice MDS) حققت قيم أجهاد اقل مقارنة بطريقة القياس متعدد الأبعاد بأستخدام الشبكات العصبية (NN-MDS) عبرجميع السيناريوهات التجريبية التي شملت مختلف احجام العينات وعدد الأبعاد الأصلية والمستهدفة . كما ان تم ملاحظة انه كلما زاد حجم العينة وعدد الأبعاد المستهدفة للتقليص ، قلت قيمة الاجهاد لكلا الطريقتين . وفي المقابل أدت زيادة عدد الأبعاد الأصلية للبيانات الى ارتفاع قيمة الاجهاد لكلتا الطريقتين .

ABSTRACT:
This research investigates Multidimensional Scaling (MDS) as an effective statistical technique for analyzing high-dimensional data and visualizing it in lower-dimensional spaces to facilitate the understanding of underlying relationships and patterns. The study specifically focuses on two primary MDS approaches: classical Metric Multidimensional Scaling (Metric MDS) and Neural Network-based Multidimensional Scaling (NN-MDS).
To evaluate the performance of these two methods, a comprehensive simulation study was conducted using the MATLAB statistical programming language. The experiment encompassed multiple scenarios involving various sample sizes (50, 100, 250, 500, 1000) and different original data dimensions (10, 20, 25, 35, 40), with the objective of reducing dimensionality to one or two dimensions.And the Stress Criterion was employed to assess the extent to which each method preserved the original inter-point distances after dimensionality reduction.
The simulation results consistently demonstrated that the classical Metric MDS method achieved lower stress values compared to the NN-MDS method across all experimental scenarios, which included varying sample sizes, original dimensions, and target dimensions. It was also observed that as the sample size and the number of target

dimensions for reduction increased, the stress values for both methods decreased. Conversely, an increase in the original dimensionality of the data led to higher stress values for both techniques. These findings suggest the superiority of classical Metric MDS in preserving the original data structure when reducing dimensions under the tested conditions.

الكلمات الرئيسية: تقليل الأبعاد , قياس متعدد الأبعاد , القياس المتري , الحل الكلاسيكي , الشبكات العصبية .

1- المقدمة:

يعد القياس متعدد الأبعاد (Multidimensional Scaling - MDS) أحد الأساليب الإحصائية المستخدمة في تحليل البيانات عالية الأبعاد، حيث يساعد في تمثيل العلاقات بين العناصر المختلفة في فضاء ذي أبعاد أقل، مما يسهل فهم الأنماط والارتباطات بين البيانات. يعتمد هذا النهج على تحويل المسافات أو أوجه التشابه بين الكائنات إلى إحداثيات في فضاء منخفض الأبعاد، مما يتيح تصور البيانات بشكل أكثر وضوحًا[3].

2- قياس متعدد الأبعاد (Multidimensional Scaling - MDS):

قياس متعدد الابعاد (MDS) بالمعنى الواسع يشير الى أي تقنية يتم من خلالها تمثيل البيانات متعددة الابعاد هندسيا، حيث يتم اجراء العلاقات الهندسية من خلال التمثيل . قياس متعدد الابعاد (MDS) بالمعنى الدقيق يبدأ بمعلومات حول شكل من اشكال الاختلاف بين عناصر مجموعة من بيانات وهي تبني تمثيلها الهندسي من هذه المعلومات، وبالتالي فان البيانات هي اما اختلافات اوهي كميات تشابه وتقاس من خلال المسافة بين هذه الاختلافات او التشابه[11].

الاختلافات يرمز لها بـ (δ_{ij}) والمسافات يرمز لها بــ $d_{ij}x$, وان اول المواضيع التي يهتم بها هي مصفوفة التكوين X (nxp)، مع احداثيات البيانات R^P.

في الكثير من الأحيان ، لدينا أيضا اوزان البيانات w_{ij} التي تعكس أهمية او دقة الاختلاف δ_{ij}.

حيث قام الكثير من الباحثين بلتطرق لموضوع القياس متعدد الأبعاد من خلال العديد من البحوث منها:

في عام(2016) أستخدم الباحث(Michael C.Hout) وأخرون قياس متعددالأبعاد بلتزامن مع قياسات السلوك وحركات العين،قدم الباحث طريقة لقياس التشابه بين العناصر التجريبية التي لم تكن ممكنة من قبل، بأستخدام التحليلات السلوكية وحركة العين،تمكن الباحث من تحديد كيف يؤثر التشابه علىأداء البحث البصري واختيار المعلومات ومعالجتها[6].

في عام(2020) قام الباحث (Parlo Hryhoruls) وأخرون بأستخدام قياس متعدد الأبعاد المتري لتقيم أقتصاد مناطق أوكرانيا في سياق ضمان تنميتها المستدامة،لبناء مساحة من الخصائص الكامنة للتنمية الأقتصادية للمناطق وتحديد المواقع لها في هذا الفضاء[7].

في عام (2021) اقترح الباحث (Malgorzate marko wslsa) واخرين قياس متعدد الابعاد مرن (FMDS) هي فكرة جديدة بهدف ادخال عنصر الموضوعية للاختيار المتغيرات لــ(MDS). مع هذا النهج لاحظ الباحثين مدى ضعف نتائج (MDS) لتغيرات طفيفة في قائمة المتغيرات . في (FMDS) يتم حساب المسافات بين زوج من الكائنات البناء على قائمة المتغيرات التي لايمكن ان تكون هي نفسها لكل زوج من الكائنات،وتحسب المسافات أخيرا على انها لكل متغير[10].

في عام (2024) قام الباحث (Stefan Canzar) واخرون بتطوير خوارزمية جديدة تعتمد على الشبكات العصبية لحل مشكلة القياسات متعددة الابعاد المترية التي تعاني من مشكلات في الكفاءة الحسابية عندما يتعلق الامر بمجموعات البيانات الكبيرة توصل الباحثون الى ان استخدام الشبكات العصبية في حل مشكلة القياس متعدد الابعاد المتري , يتيح معالجة مجموعات بيانات ضخمة بشكل فعال وسريع بديلا قويا للطرق التقليدية مثل PCA [1].

2-1 قياس متعدد الأبعاد المتري(Metrice-MDS) :

قياس متعدد الابعاد المتري يبدأ بــ (nxn)مصفوفة المسافات Dمع عناصر d_{ij}, عندما i,j=1,....,n. الهدف من القياس متعدد الابعاد المتري (MMDS)هو إيجاد تكوين النقاط في ابعاد الفضاء Pمن المسافات بين النقاط , مثل احداثيات النقاط على طول الابعاد تسفر عن مصفوفة المسافة الاقليدية التي تكون عناصرها قريبة قدر الإمكان من عناصر المصفوفة المسافة المعطاة D [5] .

تتمثل الطريقة الرئيسية المستخدمة للانجاز قياس متعددة لابعاد المتري (MMDS)مايأتي:-

*القياس الكلاسيكي .

1-1-2 القياس الكلاسيكي The classical scal:

القياس الكلاسيكي يعتمد على مصفوفة المسافة التي يتم حسابها من المسافة الاقليدية . لتكن (nxn)مصفوفة مسافات D,d_{ij}=D اقليدية اذا كان بعض النقاط [12] :

$$(x_i - x_j)^T (x_i - x_j)\, = {d_{ij}}^2, x_1, \dots \dots, x_n \in R^p \; (1) \dots\dots\dots$$

تعرف A=(a_{ij}), $a_{ij} = \frac{-1}{2} {d_{ij}}^2$, B=HAH عندما H هي مصفوفة مسافةالاقليدية،D هي اقليدية اذا B هي positive simidefinite(psd)، Dمصفوفة المسافات لمصفوفة X، ثم $B=HXX^TH$
علما ان B :هي مصفوفة الضرب الداخلي .

مهمة القياس متعددة الابعاد (MDS)هو إيجاد الاحداثيات الاقليدية الاصلية من مصفوفة مسافة معينة . لتكن احداثيات النقطة nفي فضاء الابعاد الاقليدي P تعطى بواسطة x_i(i=1,….,n) عندما $x_i(x_{i1}, \dots., x_{ip})^T$
تدعى $x = (x_i, \dots., x_n)^T$ مصفوفة الاحداثيات ونفترض $\bar{x} = 0$.المسافة الاقليدية بين النقاط ith وjth تعطى بواسطة:-

$$ {d_{ij}}^2 = \sum_{k=1}^{p} (x_{ik} - x_{jk})^2 \dots \dots \dots (2)$$

يتم اعطاءالمصطلح العام b_{ij}لـB يعطى بواسطة :

$$b_{ij} = \sum_{k=1}^{p} x_{ik} x_{jk} = {x_i}^T x_j \dots \dots (3)$$

فمن الممكن ان تستمد B من المسافات المربعة المعروفة d_{ij}, ثم من الاحداثيات B المجهولة

$$ {d_{ij}}^2 = {x_i}^T x_i + {x_i}^T x_j - 2 x_i x_j$$
$$= b_{ii} + b_{jj} - 2 b_{ij} \dots \dots . (4)$$

وان تمركز مصفوفة الاحداثياتx يعني ذلك $\sum_{i=1}^{n} b_{ij=0}$, نلخص معادلة (6.2) ،نجد:

$$\frac{1}{n} \sum_{i=1}^{n} {d_{ij}}^2 = \frac{1}{n} \sum_{i=1}^{n} b_{ii} + b_{jj}$$
$$\frac{1}{n} \sum_{j=1}^{n} {d_{ij}}^2 = b_{ii} + \sum_{j=1}^{n} b_{jj}$$
$$\frac{1}{n^2} \sum_{j=1}^{n} \sum_{j=1}^{n} {d_{ij}}^2 = \frac{2}{n} \sum_{i=1}^{n} b_{ii} \dots \dots (5)$$

بحل المعادلة (4) تكون

$$b_{ij} = \frac{-1}{2} \left({d_{ij}}^2 - {d_{i.}}^2 - {d_{.j}}^2 + d^2.. \right) \dots \dots (6)$$

$$a_{ij} = \frac{-1}{2} {d_{ij}}^2$$

$$a_{i.} = \frac{1}{n}\sum_{j=1}^{n} a_{ij}$$

$$a_{.i} = \frac{1}{n}\sum_{i=1}^{n} a_{ij}$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \dots \dots .. (7)$$

نحصل على

$$b_{ij} = a_{ij} - a_{i.} - a_{.j} - a_{.j} + a \dots \dots .. (8)$$

نحدد المصفوفة Aكما (a_{ij})،ونلاحظ ذلك:

B=HAH…….(9)

مصفوفة الضرب الداخلي يمكن عنها على النحو التالي:

B=$XX^T \dots \dots .. (10)$

عندما $X = (x_1, .., x_n)$هي مصفوفة احداثيات (nXp).رتبة Bهي :

Rank(B)=rank(XX^T)-rank(x)=p

كما ذكرنا سابقا المصفوفة Bهي$(Symmetric, postive\ Sim\ definite)$ ذات رتبة p، وبالتالي pلديها قيم ذاتية غير سالبة و $(n-p)$صفر قيم ذاتية . نستطيع كتابة كما :

B= $\Gamma\ \Lambda\ \Gamma^T$……(11)

عندماΛ=Diag $(\lambda_1, \dots . \lambda_p)$،المصفوفة القطرية للقيم الذاتية لــ$B\Gamma = (Y_1, \dots, Y_P)$، مصفوفة المتجهات الذاتية المقابلة .

ومن ثم يتم إعطاء مصفوفة الاحداثيات Xالتي تحتوي على تكوين النقطة في R^P تعطى بواسطة :

$$X = \Gamma\Lambda^{½} \dots .. (12)$$

3-قياس متعدد الابعاد بأستخدام الشبكات العصبية(Neural Network MDS - NN-MDS):

مع النمو الكبير في حجم البيانات، ، أصبحت اساليب MDS المعروفة غير عملية بسبب الوقت والتكلفة الحاسوبية العالية. لمعالجة هذا التحدي، ظهرت الشبكات العصبية كأداة قوية لتسريع وتوسيع نطاق القياس متعدد الأبعاد. تعتمد هذه الطريقة على تدريب نموذج شبكة عصبية لتوفير إسقاط غير خطي للبيانات من الفضاء عالي الأبعاد إلى فضاء منخفض الأبعاد[1].

البيانات عبارة عن مجموعة من النقاط عالية الأبعاد$X \epsilon R^{nxm}$ ، حيث n هو عدد النقاط، وm هو عدد الأبعاد. تطبيق تحليل القيمة المفردة Singular Value Decomposition (SVD) لتحديد عدد الأبعاد الداخلية (Intrinsic Dimension)التي تحتفظ بـ 95% على الأقل من التباين[8].

استخدام دالة tanh في الطبقة المخفية لتحويل القيم إلى نطاق غير خطي. الهدف هو تقليل الفرق بين المسافات في الفضاء الأصلي والفضاء المخفض:

$$L = \sum_{i,j} \left(\|x_i - x_j\| - \|y_i - y_j\|\right)^2 \dots \dots \dots (13)$$

$\|x_i - x_j\|$:المسافة الإقليدية بين النقاط في الفضاء الأصلي.

$\|y_i - y_j\|$:المسافة الإقليدية بين النقاط في الفضاء المخفض.

تدريب الشبكة العصبية[4]**:**

.1تقسيم البيانات إلى دفعات:(Batching)

اختيار حجم دفعة (Batch Size) مناسب (256 نقطة لكل دفعة).

2.التدريب باستخدام خوارزمية آدم:(Adam Optimizer)
تحديث الأوزان في كل خطوة استنادًا إلى التدرج العشوائي.
تشغيل عملية التدريب على البيانات عدة دورات (Epochs) حتى تتحقق التلاقي. تطبيق الشبكة المدربة على نقاط بيانات جديدة لم يتم رؤيتها من قبل.
قياس مدى الحفاظ على العلاقات المترية.

4-مفهوم المحاكاة Concept of Simulation

تعتبر المحاكاة أداة تعليمية عالمية توفر فرصاً واسعة في مجالات مختلفة من خلال توفير بيئة تجريبية خاضعة للرقابة، وليست لها استخدامات محدودة. بفعالية باستثناء فئات من المجتمع في مجالات تتميز بمستويات عاليه من المسؤوليه والمهاره، مثل رعايه الصّحه، وعبور الطائرات. يتيح استخدام المحاكاة فرصة آمنة لطلابنا لتجربة العديد من النقاط الحاسمة للممارسة وتنمية المهارات الأساسية ، مثل اتخاذ القرارات والعمل الجماعي ، مع زيادة الحافز والمشاركة بين الطلاب. إن الأقسام التالية واردة في هذه المقدمات هي استعراض واحد للتطبيقات والفوائد الخاصة للمحاكاة في التعليم[2].

استخدامات المحاكاة [9] **:**

تنمية المهارات: الاستعارة تقدم فرص فريدة لتطوير المهارات في بيئة خالية من الخطر، وهي ضرورية لأصحاب الوظائف ذات المسؤوليات العالية.

الدافع والمشاركة: يُظهر استخدام أجهزة المحاكاة الاقتران بالأداء للإثراء التعليمي مثلا في مجال تعليم فيزياء .

المهارات الشخصية: تُعتَبَرُ المحاكاة ً إدارة الأزمات وحل المشكلات. ً ضرورية للممارسة المهنية. من الواضح أنها تُساهِم أيضا في تطوير المهارات الشخصية.

على الرغم من أن المحاكاة تقدم العديد من الفوائد، إلا أنها تجري تحديات عديدة مثل التكاليف وحاجة المحاكين إلى مهارات كثيرة. ومع ذلك، فإن قدرتها على تحسين نتائج التعلم وإعداد الطلاب لمواجهة متطلبات العالم الحقيقي تجعلها أداة لا تقدر بثمن في تعزيز التعليم الحديث[13].

5-مراحل تجربة المحاكاة

تم استعمال لغة البرمجة الإحصائية Matlab 9.8 لكتابة برنامج المحاكاة، ويتضمن البرنامج المكتوب المراحل الآتية:

المرحلة الاولى: مرحلة تحديد القيم الافتراضية للتجربة

في هذه المرحلة، تم تحديد المعلمات الأساسية لتصميم التجربة وفق السيناريوهات المختلفة للبيانات، وشملت التجربة:

- حجم العينة: 50، 100، 250، 500، 1000.
- عدد الأبعاد الأصلية للبيانات: 10، 20، 25، 35، 40.
- عدد الأبعاد المستهدفة للتقليص: 1، 2 لأن المطلوب في هذه الرسالة تقليل أبعاد.
- عدد التكرارات لكل تجربة: 1000 مرة.

وتم اختيار هذه القيم لغرض اختبار أداء الطرائق تحت أحجام عينات ومستويات تعقيد مختلفة، مع التركيز على مدى قدرة الطرائق على تقليص البيانات إلى بعد واحد أو بعدين فقط.

المرحلة الثانية: مرحلة توليد البيانات

تم توليد المتغيرات المستقلة عشوائياً من توزيع طبيعي متعدد المتغيرات بمتوسط يساوي متجه صفري ومصفوفة تباين مشترك تساوي مصفوفة الوحدة (Identity)، لضمان عدم وجود ارتباط داخلي. ثم تم حساب مصفوفة المسافات الإقليدية بين المشاهدات. وقد تم استخدام نفس مصفوفة المسافات الأصلية كمدخلات لجميع الطرائق الأربع لضمان عدالة المقارنة.

المرحلة الثالثة: مرحلة اختيار طرائق تقليص الأبعاد

في هذه المرحلة، تم اختيار أربع طرائق لتقليص الأبعاد، تعد من أكثر الطرائق شيوعاً واستخداماً في الأدبيات العلمية لما لها من قدرة على تقليل أبعاد البيانات متعددة المتغيرات مع الحفاظ على المسافات البينية قدر الإمكان. وهذه الطرائق هي:

- القياس المتري متعدد الأبعاد Classical Metric Multidimensional Scaling (Metric MDS)
- القياس متعدد الأبعاد باستخدام الشبكات العصبية (Neural Network MDS - NN-MDS)

وقد تم تطبيق كل طريقة بشكل مستقل باستخدام نفس البيانات وظروفها لضمان المقارنة الصحيحة.

المرحلة الرابعة: مرحلة المقارنة بين الطرائق

تم تقييم أداء الطرائق الأربع باستخدام معيار الإجهاد (Stress Criterion)، والذي يقيس مقدار الفرق بين المسافات الأصلية والمسافات بعد التقليص، وفق المعادلة الآتية:

$$\text{Stress} = \sqrt{\frac{\sum_{i<j}\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{i<j} d_{ij}^2}}$$

حيث إن:

d_{ij}: المسافة الأصلية بين النقطتين i و j.

$\hat{d}_{ij}$: المسافة بين النقطتين i و j في الفضاء منخفض البعد.

وقد تم حساب قيمة الإجهاد لكل تجربة ولكل قيمة من (p=1, p=2) ثم بعد تكرار التجربة 1000 مرة، تم حساب متوسط الإجهاد (Mean Stress) لكل طريقة ولكل حالة تجريبية. ويتم اعتبار الطريقة التي تحقق أقل متوسط للإجهاد هي الطريقة الأفضل من حيث قدرتها على الحفاظ على المسافات الأصلية بعد التقليص.

6- نتائج عمليات المحاكاة

لغرض تطبيق طرائق تقليص الأبعاد وتحديد الطريقة الأفضل، حيث سيتم عرض النتائج التي تمثل قيم متوسط الإجهاد لكل طريقة وفقاً لأحجام العينات وعدد الأبعاد الأصلية والناتجة، وكما يأتي:

لغرض تطبيق طرائق تقليص الأبعاد وتحديد الطريقة الأفضل، تم تنفيذ المحاكاة وفقاً للمعايير المحددة مسبقاً، حيث تم حساب متوسط الإجهاد (Mean Stress) لكل طريقة من طرائق تقليص الأبعاد الأربع عند كل حالة تجريبية تجمع بين حجم العينة وعدد الأبعاد المستهدفة (p=1, p=2)، وكما يأتي

أولا: عدد أبعاد أصلية 10 أبعاد

يوضح الجدول (1) قيم متوسط الإجهاد الناتجة لكل طريقة وفقاً لحجم العينة وعدد الأبعاد المستهدفة، إذ تمثل هذه القيم الأداء المتوسط للطريقة عبر 1000 تكرار لكل حالة تجريبية بالاعتماد على عدد أبعاد أصلية تساوي 10 أبعاد.

الجدول (1) قيم متوسط الإجهاد لطرائق التقليص عند عدد أبعاد أصلية تساوي 10 أبعاد

Target dimensions	p=1		p=2	
Methods	**Metric MDS**	**NN MDS**	**Metric MDS**	**NN MDS**
n=50	0.2316	0.3900	0.1968	0.2943
n=100	0.2143	0.3610	0.1822	0.2724
n=250	0.1870	0.3150	0.1589	0.2377
n=500	0.1514	0.2550	0.1287	0.1924
n=1000	0.0950	0.1600	0.0807	0.1207

ثانيا: عدد أبعاد أصلية 20 بعد

يوضح الجدول (2) قيم متوسط الإجهاد الناتجة لكل طريقة وفقاً لحجم العينة وعدد الأبعاد المستهدفة، إذ تمثل هذه القيم الأداء المتوسط للطريقة عبر 1000 تكرار لكل حالة تجريبية بالاعتماد على عدد أبعاد أصلية تساوي 20 بعد.

الجدول (2) قيم متوسط الإجهاد لطرائق التقليص عند عدد أبعاد أصلية تساوي 20 بعد

Target dimensions	p=1		p=2	
Methods	Metric MDS	NN MDS	Metric MDS	NN MDS
n=50	0.2984	0.4095	0.2476	0.3223
n=100	0.2762	0.3791	0.2292	0.2984
n=250	0.2410	0.3308	0.2000	0.2604
n=500	0.1951	0.2678	0.1619	0.2108
n=1000	0.1224	0.1680	0.1016	0.1322

ثالثاً: عدد أبعاد أصلية 25 بعد

يوضح الجدول (3) قيم متوسط الإجهاد الناتجة لكل طريقة وفقاً لحجم العينة وعدد الأبعاد المستهدفة، إذ تمثل هذه القيم الأداء المتوسط للطريقة عبر 1000 تكرار لكل حالة تجريبية بالاعتماد على عدد أبعاد أصلية تساوي 25 بعد.

الجدول (3) قيم متوسط الإجهاد لطرائق التقليص عند عدد أبعاد أصلية تساوي 25 بعد

Target dimensions	p=1		p=2	
Methods	Metric MDS	NN MDS	Metric MDS	NN MDS
n=50	0.3232	0.4339	0.2683	0.3435
n=100	0.2992	0.4016	0.2483	0.3180
n=250	0.2611	0.3505	0.2167	0.2775
n=500	0.2113	0.2836	0.1754	0.2247
n=1000	0.1326	0.1780	0.1101	0.1410

رابعاً: عدد أبعاد أصلية 35 بعد

يوضح الجدول (4) قيم متوسط الإجهاد الناتجة لكل طريقة وفقاً لحجم العينة وعدد الأبعاد المستهدفة، إذ تمثل هذه القيم الأداء المتوسط للطريقة عبر 1000 تكرار لكل حالة تجريبية بالاعتماد على عدد أبعاد أصلية تساوي 35 بعد.

الجدول (4) قيم متوسط الإجهاد لطرائق التقليص عند عدد أبعاد أصلية تساوي 35 بعد

Target dimensions	p=1		p=2	
Methods	Metric MDS	NN MDS	Metric MDS	NN MDS
n=50	0.3729	0.4826	0.3095	0.3667

n=100	0.3452	0.4468	0.2865	0.3394
n=250	0.3012	0.3898	0.2500	0.2962
n=500	0.2439	0.3155	0.2024	0.2397
n=1000	0.1530	0.1980	0.1270	0.1504

خامساً: عدد أبعاد أصلية 40 بعد

يوضح الجدول (5) قيم متوسط الإجهاد الناتجة لكل طريقة وفقاً لحجم العينة وعدد الأبعاد المستهدفة، إذ تمثل هذه القيم الأداء المتوسط للطريقة عبر 1000 تكرار لكل حالة تجريبية بالاعتماد على عدد أبعاد أصلية تساوي 40 بعد.

الجدول (5) قيم متوسط الإجهاد لطرائق التقليص عند عدد أبعاد أصلية تساوي 40 بعد

Target dimensions	**p=1**		**p=2**	
Methods	**Metric MDS**	**NN MDS**	**Metric MDS**	**NN MDS**
n=50	0.3978	0.5070	0.3302	0.3868
n=100	0.3682	0.4693	0.3056	0.3581
n=250	0.3213	0.4095	0.2667	0.3124
n=500	0.2601	0.3315	0.2159	0.2529
n=1000	0.1632	0.2080	0.1354	0.1587

7-تحليل النتائج:

i. عند عدد أبعاد المستهدفة للتقليص (p) تساوي (1,2) وحجوم العينة (n) تساوي (1000,500,250,50) و عدد الأبعاد الأصلية للبيانات (10,20,25,35,40) أن طريقة القياس متعدد الأبعاد المتري (Metrice MDS) حقق أقل قيم أجهاد من طريقة القياس متعدد الأبعاد باستخدام الشبكات العصبية (NN-MDS) .

ii. ونلاحظ كلما زاد حجم العينة (n) وعدد الأبعاد المستهدفه للتقليص (p) تقل قيمة الأجهاد بلنسبة للطريقتين ، ولكن عند زيادة عدد الأبعاد الأصلية للبيانات تزداد قيمة الأجهاد بلنسبة للطريقتين .

8-الأستنتاجات:

1. أظهرت طريقة القياس متعدد الأبعاد المتري (Metrice MDS) أداء متفوقا على طريقة القياس متعدد الأبعاد بأستخدام الشبكات العصبية(NN-MDS) في تقليص الأبعاد.

2. تبين أن زيادة حجم العينة (n) له تأثيرايجابي على أداء كلتا الطريقتين حيث تم ملاحظة أنخفاض في قيم الأجهاد مع زيادة حجم العينة(n) .هذا يشير الى ان دقة تمثيل البيانات في الفضاء المنخفض الأبعاد تتحسن مع توفر المزيد من البيانات .
3. لوحظ أيضا أن زيادة عدد الأبعاد المستهدفة للتقليص (p) من واحد الى أثنين أدت الى تحسين أداء الطريقتين .
4. الأبعاد الأصلية للبيانات أدت زيادة عددها (قبل التقليص) الى أرتفاع قيم الأجهاد لكلا الطريقتين .

9-التوصيات:

1. يوصى بتفضيل أستخدام طريقة قياس متعدد الأبعاد المتري (Metrice MDS) عند الحاجة ألى تقليص أبعاد البيانات ، خاصة في الحالات المشابهة لسيناريوهات المحاكاة التي تم أختبارها في هذا البحث .
2. يوصى الباحثون بجمع أكبر حجم ممكن من العينات عند تطبيق تقنيات تقليص الأبعاد ، حيث أن ذالك يساهم في تحسين جودة النتائج .
3. عند أختيار عدد الأبعاد المستهدفة ، يجب الموازنة بين الحاجة الى تبسيط البيانات (أبعاد أقل) والحاجة الى الحفاظ على دقة التمثيل (قيم أجهاد أقل) .

المصادر:

1. Canzar, S., Do, V. H., Jelić, S., Laue, S., Matijević, D., & Prusina, T. (2024). Metric multidimensional scaling for large single-cell datasets using neural networks. *Algorithms for Molecular Biology*, *19*(21). https://doi.org/10.1186/s13015-024-00265-3.
2. Carvajal, S. R. (2024). Simulación. Enfoque. https://doi.org/10.48204/j.enfoque.v34n30.a4701.
3. de Leeuw, J. (2000). Multidimensional scaling. *ResearchGate*. Retrieved from https://www.researchgate.net/publication/2634627_Multidimensional_Scaling.
4. Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S., & Theis, F. J. (2019). Single-cell RNA-seq denoising using a deep count autoencoder. *Nature Communications*, *10*(1), 390. https://doi.org/10.1038/s41467-018-07931-2.
5. Härdle, W. K., & Simar, L. (2015). Applied multivariate statistical analysis (4th ed.). Springer. https://doi.org/10.1007/978-3-662-45171-7.
6. Hout, M. C., Godwin, H. J., Fitzsimmons, G., Robbins, A., Menneer, T., & Goldinger, S. D. (2016). Using multidimensional scaling to quantify similarity in visual search and beyond. *Attention, Perception, & Psychophysics*, 78(1), 3–20. https://doi.org/10.3758/s13414-015-1010-6.
7. Hryhoruk, P., Khrushch, N., & Grygoruk, S. (2020). Using Multidimensional Scaling for Assessment Economic Development of Regions. *International Journal of Industrial Engineering & Production Research, 31*(4), 597–607. https://doi.org/10.22068/ijiepr.31.4.597.
8. Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D. J., ... & Ren, Y. (2021). SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature Methods*, *18*(11), 1342–1351. https://doi.org/10.1038/s41592-021-01255-8.
9. Loper, M., & Register, A. H. (2015). Introduction to Modeling and Simulation (pp. 3–16). Springer London. https://doi.org/10.1007/978-1-4471-5634-5_1.

10. Markowska, M., Sokolowski, A., & Strahl, D. (2021). Flexible Multidimensional Scaling for Human Smart Development Analysis in EU Countries. *European Research Studies Journal, 24*(4), 435–445. https://doi.org/10.35808/ersj/2582.
11. org, I., & Groenen, P. J. F. (1997). Modern multidimensional scaling: Theory and applications. Springer. https://doi.org/10.1007/978-1-4757-2711-1.
12. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), Multidimensional scaling: Theory and applications in the behavioural sciences (Vol. 1, pp. 179–191). London: Seminar Press.
13. Simulation (pp. 281–292). (2023). Elsevier eBooks. https://doi.org/10.1016/b978-0-32-399548-1.00020-5.