Prep Course Project Presentation

OUTLIER DETECTION AND ROBUST PCA USING A CONVEX MEASURE OF INNOVATION

Authors : Mostafa Rahmani, Ping Li

COURSE DETAILS

Course Title : **IE 506 : Machine Learning: Principles and Techniques**
Instructor : **Prof. P Balamurugan**

THIS WORK IS DONE AS PART OF IE 506 COURSE PROJECT

TEAM DETAILS

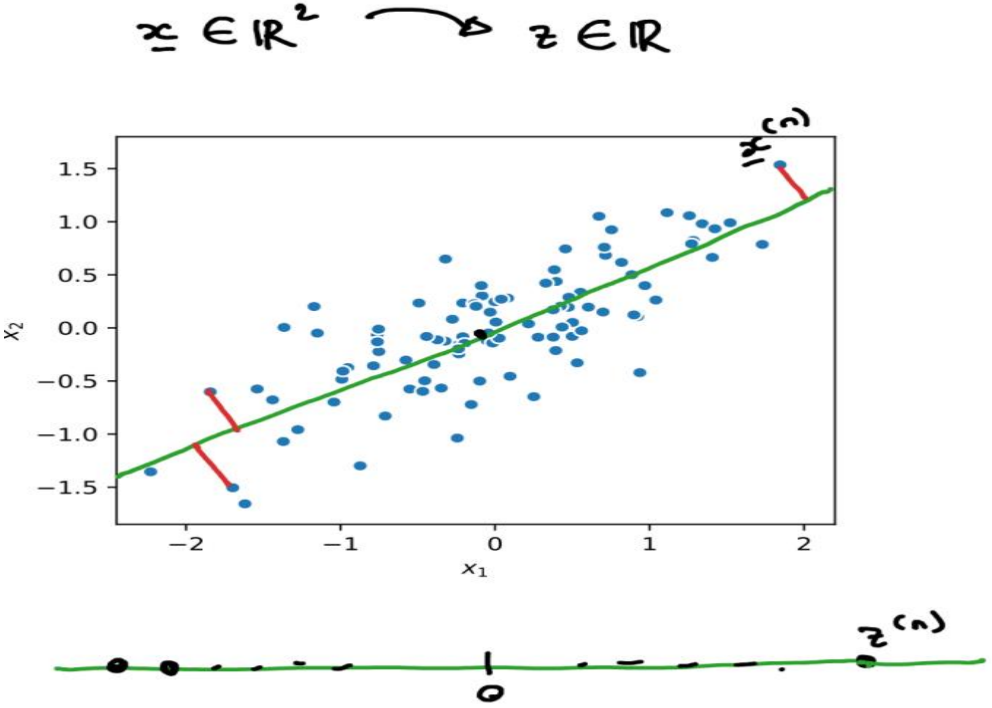
Team : **MLTorch**
Member : **Ashish Kumar Uchadiya 22M1521**
Member : **Akansh Verma 22M1515**
TA Incharge : **Krushna Salunke & Vivek Seth**

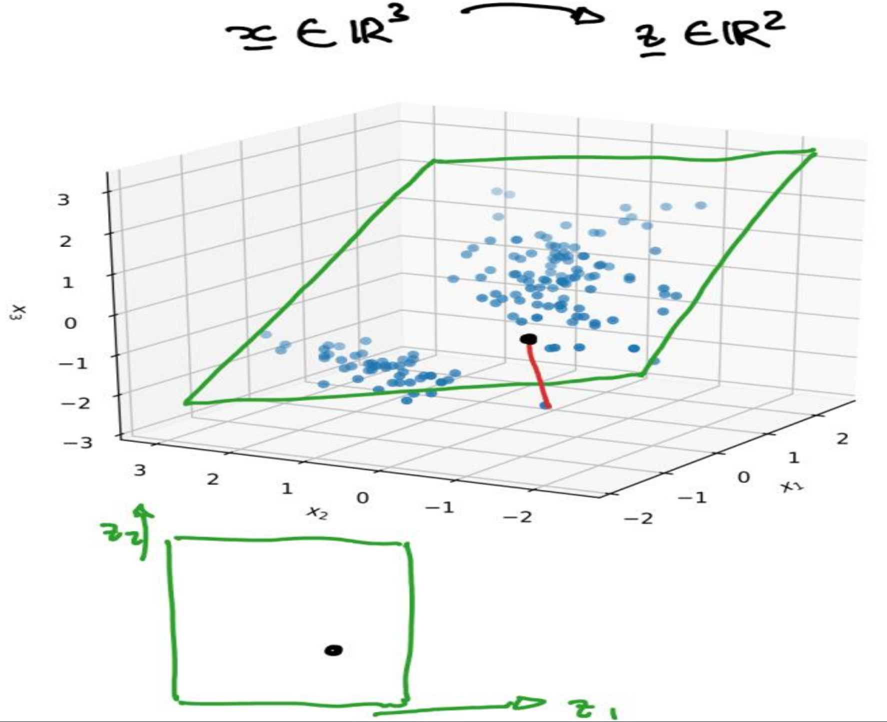
01000100 01000101 01000101 01010000 00100000 01001100 01000101 01000001 01010010 01001110 01001001 01001110 01000111 00001010

</ Presentation Outline />

</ PCA PRINCIPAL COMPONENTS ANALYSIS />

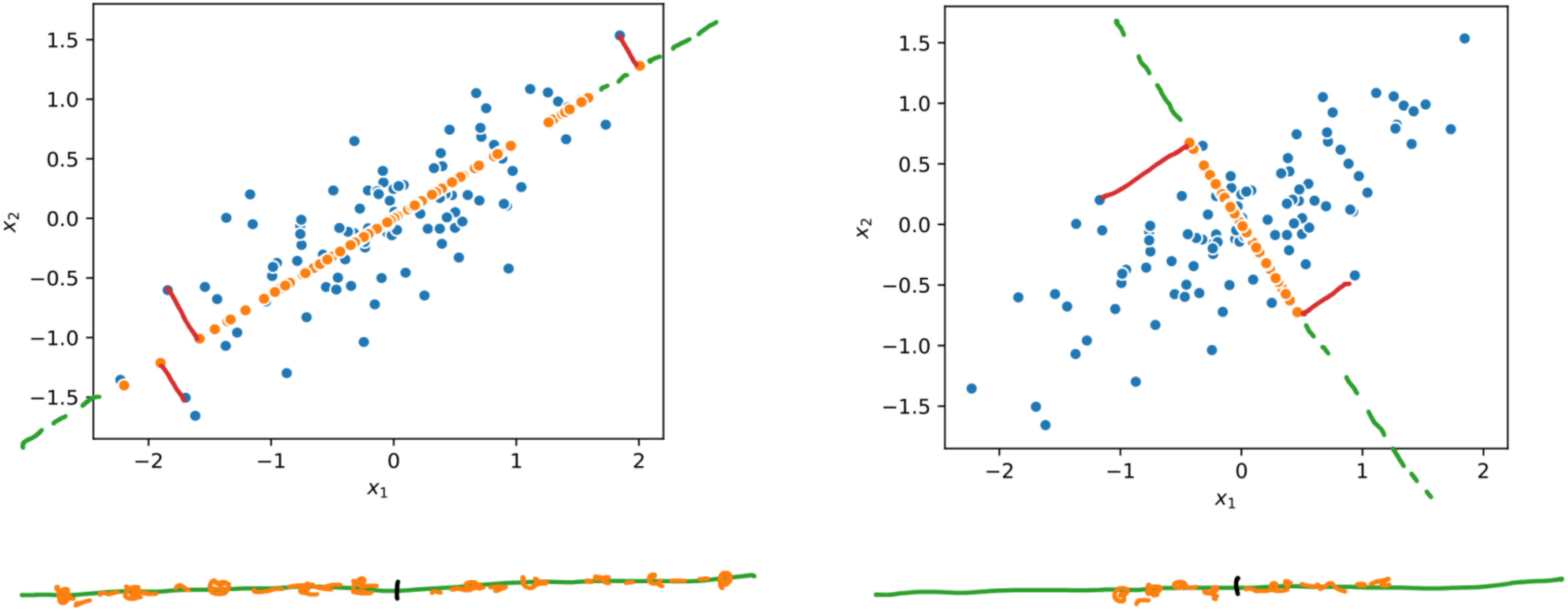
Linear projection is a fundamental concept in Principal Component Analysis (PCA), a technique used for dimensionality reduction.





</ PCA **AIM** />

PCA aims to transform high-dimensional data into a lower-dimensional subspace while retaining as much variance as possible.



$\widehat{\sigma_1} = 0.9493$ $\widehat{\sigma_2} < \widehat{\sigma_1}$ $\widehat{\sigma_2} = 0.1017$

LAGRANGE MULTIPLIER

Want to optimise $f(x)$,

subject to some constraint $g(x) = 0$.

Then we define a new objective:

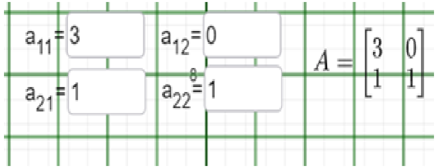
$$\mathbf{J(x,\ \lambda) = f(x) + \lambda\, g(x)}$$

and optimise w.r.t. both x and λ.

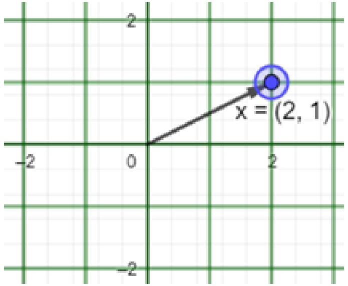
</ EIGENVALUES AND EIGENVECTORS />

An eigenvector, is a special kind of arrow in linear algebra that remains pointed in the same direction after undergoing a specific transformation represented by a matrix. The only change it might experience is being stretched or shrunk by a certain factor called the eigenvalue.

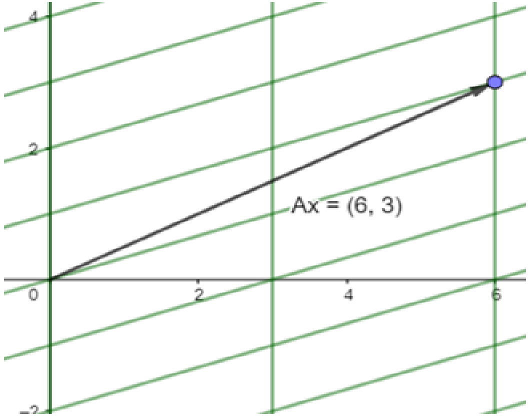
$$Ax = \lambda X$$



x



=



Linear transformer
A

Eigen Vector
x

After transformation
λX

</ PCA **GEOMETRIC INTUITION** />

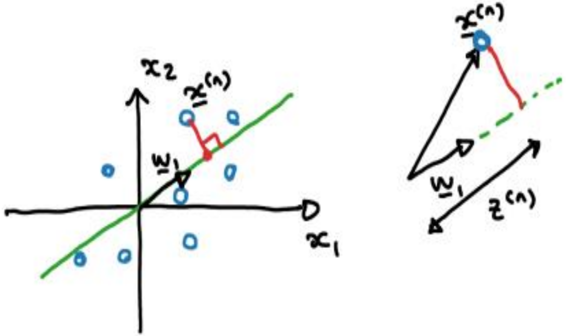
We want to project $\underline{x}^n \in \mathrm{R}^\mathrm{d}$ to $\underline{Z}^n \in R^m$ with M<D.
(Normally assume data have been normalised to have zero-mean.)

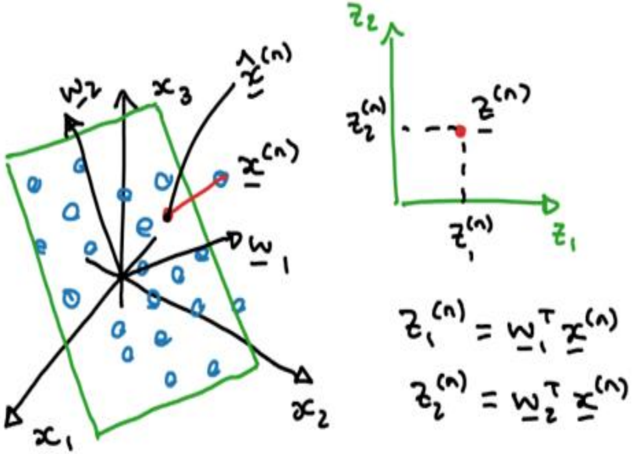
Use M "projection vectors" $\underline{W}_m \in \mathrm{R}^\mathrm{d}$

Projection vectors $\underline{W}_1$,……., $\underline{W}_m$ are unit length and orthogonal, i.e. $\|\underline{W}_m\|=1$ and $\underline{W}_i^\top \underline{W}_j = 0 \quad \forall\, i \neq j$

The projection of the n^{th} item $\underline{x}^n$ onto the m^{th} dimension is

$$\mathbf{Z}_m^{(n)} = \underline{w}_m^\top \underline{x}^{(n)}$$





</ PCA **PROJECTION VECTORS** />

So $\underline{x}^{(n)}$ is mapped to

$$\underset{M\times 1}{\underline{z}^{(n)}} = \begin{bmatrix} z_1^{(n)} \\ z_2^{(n)} \\ \vdots \\ z_M^{(n)} \end{bmatrix} = \begin{bmatrix} \underline{w}_1^T \underline{x}^{(n)} \\ \underline{w}_2^T \underline{x}^{(n)} \\ \vdots \\ \underline{w}_M^T \underline{x}^{(n)} \end{bmatrix}$$

$$= \underset{M\times D}{\begin{bmatrix} — \underline{w}_1^T — \\ — \underline{w}_2^T — \\ \vdots \\ — \underline{w}_M^T — \end{bmatrix}} \underset{D\times 1}{\begin{bmatrix} x_1^{(n)} \\ x_2^{(n)} \\ \vdots \\ x_D^{(n)} \end{bmatrix}} = \underline{W}^T \underline{x}^{(n)}$$

$$\underline{W} = \begin{array}{cc} & M \\ D & \underset{D\times M}{\begin{bmatrix} | & | & & | \\ \underline{w}_1 & \underline{w}_2 & \cdots & \underline{w}_M \\ | & | & & | \end{bmatrix}} \end{array}$$

</ PCA LEARNING PROJECTION VECTORS />

Data $x^{(1)}, \ldots, \underline{x}^{(N)}$ have been mean normalized(zero mean)

We want to find $\underline{w}_1, \ldots, \underline{w}_M$

$$\|\underline{w}_m\| = 1 \quad \forall m$$

$$\underline{w}_i^\top \underline{w}_j = 0 \quad \forall i \neq j$$

Problem: $\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_M$

Wants to find

So that (sample) variance is maximized

$$\begin{aligned}
\hat{\sigma}^2_{z_1} &= \frac{1}{N}\sum_{n=1}^{N}\left(z_1^{(n)} - \bar{z}\right)^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(z_1^{(n)}\right)^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(\underline{w}_1^\top \underline{x}^{(n)}\right)^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(\underline{w}_1^\top \underline{x}^{(n)}\right)\left(\underline{w}_1^\top \underline{x}^{(n)}\right)^\top \\
&= \underline{w}_1^\top \left[\frac{1}{N}\sum_{n=1}^{N}\underline{x}^{(n)}\left(\underline{x}^{(n)}\right)^\top\right]\underline{w}_1 \\
&= \underline{w}_1^\top \hat{\Sigma}\underline{w}_1
\end{aligned}$$

Covariance matrix

</ PCA **RELATIONSHIP TO SVD** />

Decomposition method that take one matrix x and decompose it into 3 other matrix that has some useful properties

X = U S V^T
X NXN   S NXD V DXD
U columns are orthonormal i.e $U^TU = I$
S Columns are orthonormal

Singular value Decomposition:

$$\underset{\text{NXD}}{\underline{X}} = \underset{\text{NXN}}{\underline{U}}\ \underset{\text{NXD}}{\underline{S}}\ \underset{\text{DXD}}{\underline{V}^\top}$$

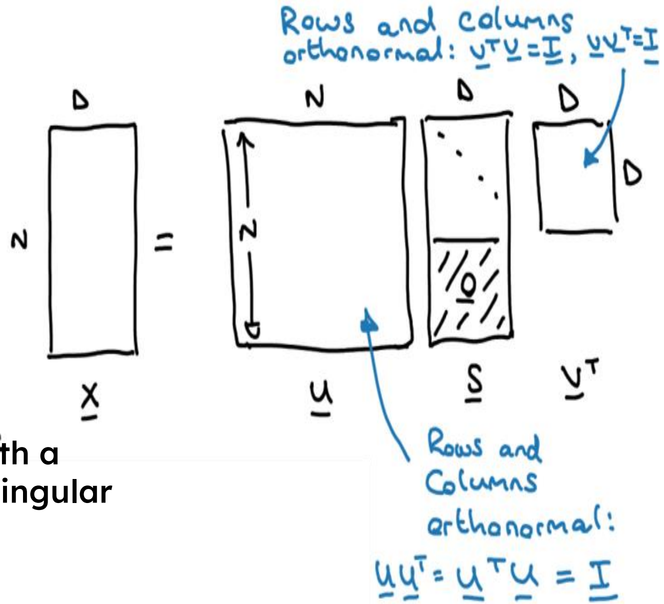
Relationship to PCA:
Take SVD of the design matrix x:

$$\underline{X} = \underline{U}\ \underline{S}\ \underline{V}^\top$$

Then
$$\begin{aligned} \underline{X}^\top \underline{X} &= \underline{V}\ \underline{S}^\top\ \underline{U}^\top\ \underline{U}\ \underline{S}\ \underline{V}^\top \\ &= \underline{V}\ \underline{S}^\top\ \underline{S}\ \underline{V}^\top \\ &= \underline{V}\ \underline{D}\ \underline{V}^\top \end{aligned}$$

$$\underline{D} = \underline{S}^\top\ \underline{S}$$

Diagonal with a squares of singular values



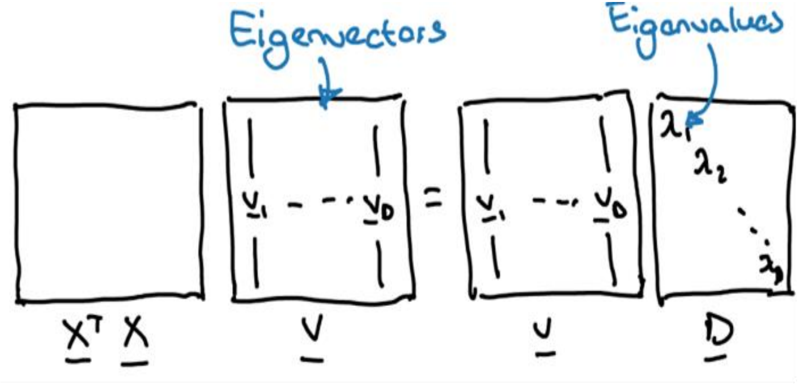
</ PCA **RELATIONSHIP TO SVD** />

$$\underline{X}^\top \underline{X} = \underline{V}\, \underline{D}\, \underline{V}^\top$$

Right multiplying by V both the sides

$$(\underline{X}^\top \underline{X})\, \underline{V} = \underline{V}\, \underline{D}$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} \underline{x}^{(n)} \left(\underline{x}^{(n)}\right)^\top$$



</ PCA AS A CONVEX MAXIMIZATION PROBLEM />

Want to maximise $\hat{\sigma}^2_{z_1}$ subject to

$$\|\underline{w}_1\|^2 = 1, \text{ i.e. } \underline{w}_1^T \underline{w}_1 = 1$$

Loss (minimise)

Use Lagrange multiplier:

$$J(\underline{w}_1) = -\hat{\sigma}^2_{z_1} + \lambda(\underline{w}_1^T \underline{w}_1 - 1)$$

$$= -\underline{w}_1^T \hat{\underline{\Sigma}} \underline{w}_1 + \lambda(\underline{w}_1^T \underline{w}_1 - 1)$$

Minimise w.r.t. $\underline{w}_1$:

$$\frac{\partial J(\underline{w}_1)}{\partial \underline{w}_1} = -\cancel{2}\hat{\underline{\Sigma}} \underline{w}_1 + \cancel{2}\lambda \underline{w}_1 = \underline{0}$$

$$\underbrace{\hat{\underline{\Sigma}} \underline{w}_1 = \lambda \underline{w}_1}_{\text{Eigenvalue / eigenvector equation}} \quad \dots (1)$$

Which eigenvector/value do we use?

From (1): $\underline{w}_1^T \hat{\underline{\Sigma}} \underline{w}_1 = \lambda \underline{w}_1^T \underline{w}_1$

$$\underline{w}_1^T \hat{\underline{\Sigma}} \underline{w}_1 = \lambda$$

Want this maximised $\rightarrow$ $\hat{\sigma}^2_{z_1} = \lambda$

So pick eigenvector corresponding to largest eigenvalue.

How do we find $\underline{w}_2$, with $\|\underline{w}_2\|^2 = 1$ and $\underline{w}_1^T \underline{w}_2 = 0$?

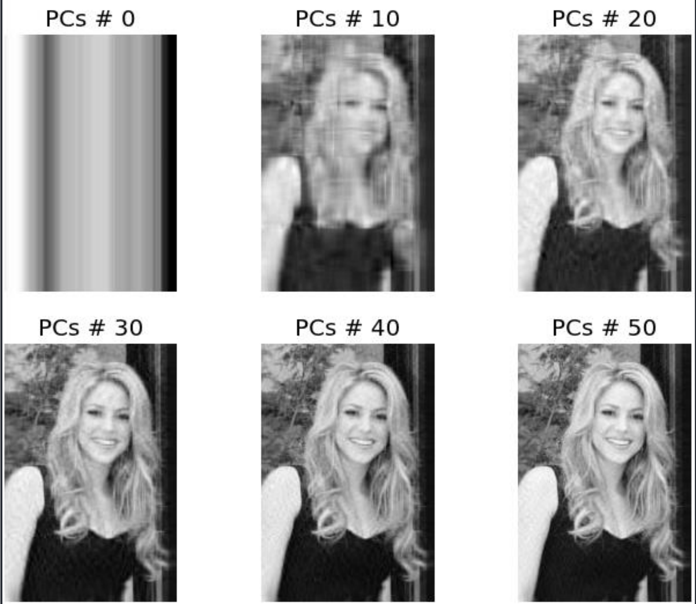
Repeat above steps:

$$\hat{\underline{\Sigma}} \underline{w}_2 = \lambda_2 \underline{w}_2$$

Pick eigenvector corresponding to 2nd highest eigenvalue, etc.

</ PCA REAL WORLD APPLICATION />

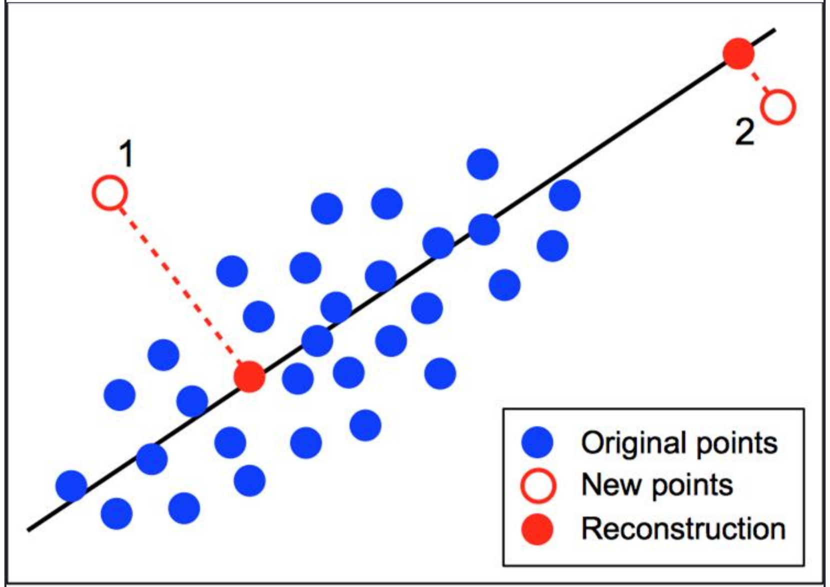
IMAGE COMPRESSION



We can see that 40 principal components are enough to reconstruct the original image

Image source: PCA and image compression with numpy

ANOMALY DETECTION



By identifying deviations from the principal components, PCA can help detect unusual data points that might indicate fraud, system failures, or other anomalies

Image source: Anomaly detection using PCA reconstruction error

</ PCA

PCA LIMITATIONS AND RPCA

/>

1. Information loss
2. PCA assumes Linear Relationships between variables
3. Outliers can significantly influence the direction of the principal components
4. PCA is sensitive to the scale of your variables

</ PCA **CODE, DATASET** />

CODE: **NOT AVAILABLE** Searched on *paperwithcode.com*

DATASET: **HOPKINS 155**

https://paperswithcode.com/dataset/hopkins155

Introduced by Roberto Tron et al. in

A Benchmark for the Comparison of 3-D Motion Segmentation Algorithms

THANK YOU

</ REFERENCES />

Paper References

[1] PAPER : **Outlier Detection and Robust PCA Using a Convex Measure of Innovation, NeurIPS 2019**
| Authors : Mostafa Rahmani, Ping Li | **Link :** http://papers.nips.cc/paper/9568-outlier-detection-and-robust-pca-using-a-convex-measure-of-innovation.pdf

[2] PAPER : **Innovation Pursuit: A New Approach to the Subspace Clustering Problem, ICML 2017**
| Authors : Mostafa Rahmani, George Atia | **Link :** http://proceedings.mlr.press/v70/rahmani17b/rahmani17b.pdf

[3] PAPER : **Coherence Pursuit: Fast, Simple, and Robust Subspace Recovery, ICML 2017**
| Authors : Mostafa Rahmani, George Atia | **Link :** http://proceedings.mlr.press/v70/rahmani17a/rahmani17a.pdf

[4] PAPER : **Outlier Detection and Data Clustering via Innovation Search, 30 Dec 2019**
| Authors : Mostafa Rahmani, Ping Li | **Link :** https://arxiv.org/pdf/1912.12988v1.pdf

[5] PAPER : **Outlier Detection and Data Clustering via Innovation Search, 30 Dec 2019**
| Authors : Mostafa Rahmani, George Atia | **Link :** https://arxiv.org/pdf/1912.12988v1.pdf

Article References

[1] ARTICLE : **Eigen decomposition of a covariance matrix**
| Editor : Vincent Spruyt | **Link :** https://www.visiondummy.com/2014/04/geometric-interpretation-covariance-matrix/#Eigendecomposition_of_a_covariance_matrix

[2] ARTICLE : **PCA and image compression with numpy**
| Editor : The Glowing Python | **Link :** https://glowingpython.blogspot.com/2011/07/pca-and-image-compression-with-numpy.html

[3] ARTICLE : **Anomaly detection using PCA reconstruction error**
| Editor : StackExchange | **Link :** https://stats.stackexchange.com/questions/259806/anomaly-detection-using-pca-reconstruction-error

[4] ARTICLE : **DatA414 Introduction to machine learning**
| Editor : Herman Kamper | **Link :** https://www.kamperh.com/data414/

</ REFERENCES />

Video Reference

[1] VIDEO : **Principal Component Analysis (PCA) _ Part 1 _ Geometric Intuition**
| Creator : Nitish Singh | **Link :** https://youtu.be/iRbsBi5W0-c?si=HMIw7VAcwwptB27l

[2] VIDEO : **Principal Component Analysis (PCA) | Part 2 | Problem Formulation and Step by Step Solution**
| Creator : Nitish Singh | **Link :** https://www.youtube.com/watch?v=tXXnxjj2wM4

[3] VIDEO : **Principal Component Analysis (PCA) | Part 3 | Code Example and Visualization**
| Creator : Nitish Singh | **Link :** https://www.youtube.com/watch?v=tofVCUDrg4M

[4] VIDEO : **Robust Principal Component Analysis (RPCA)**
| Creator : Steve Brunton | **Link :** https://www.youtube.com/watch?v=yDpz0PqULXQ&t=21s

[5] VIDEO : **PCA 1 - Introduction**
| Creator : Herman Kamper | **Link :** https://www.youtube.com/playlist?list=PLmZlBIcArwhMfNuMBg4XR-YQ0QIqdHCrl

Image References

[1] https://glowingpython.blogspot.com/2011/07/pca-and-image-compression-with-numpy.html

[2] https://i.stack.imgur.com/1j5X1.png