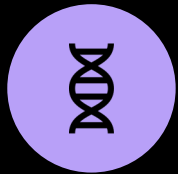




PRINCIPAL COMPONENT ANALYSIS

PRINCIPAL COMPONENT ANALYSIS



What is PCA?



Benefit of PCA



Calculation
step



Discussion
about PCA



Code



Further reading

What is PCA?

eigenvalues = 20, 5, 2, 0.5, 0.1, 0.03, 0.006

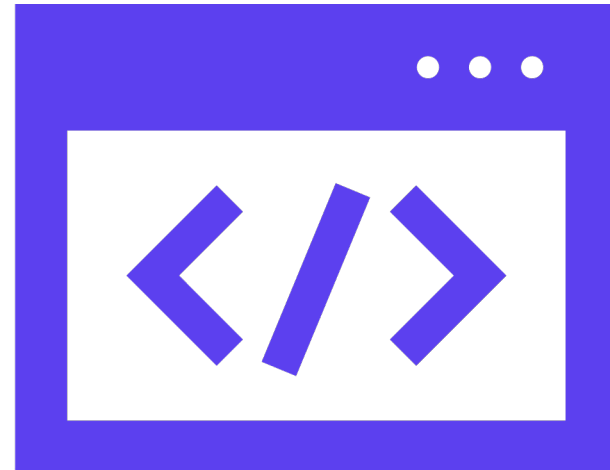
2 มิติ รักษา information ได้ 90.46%

3 มิติ รักษา information ได้ 97.69%

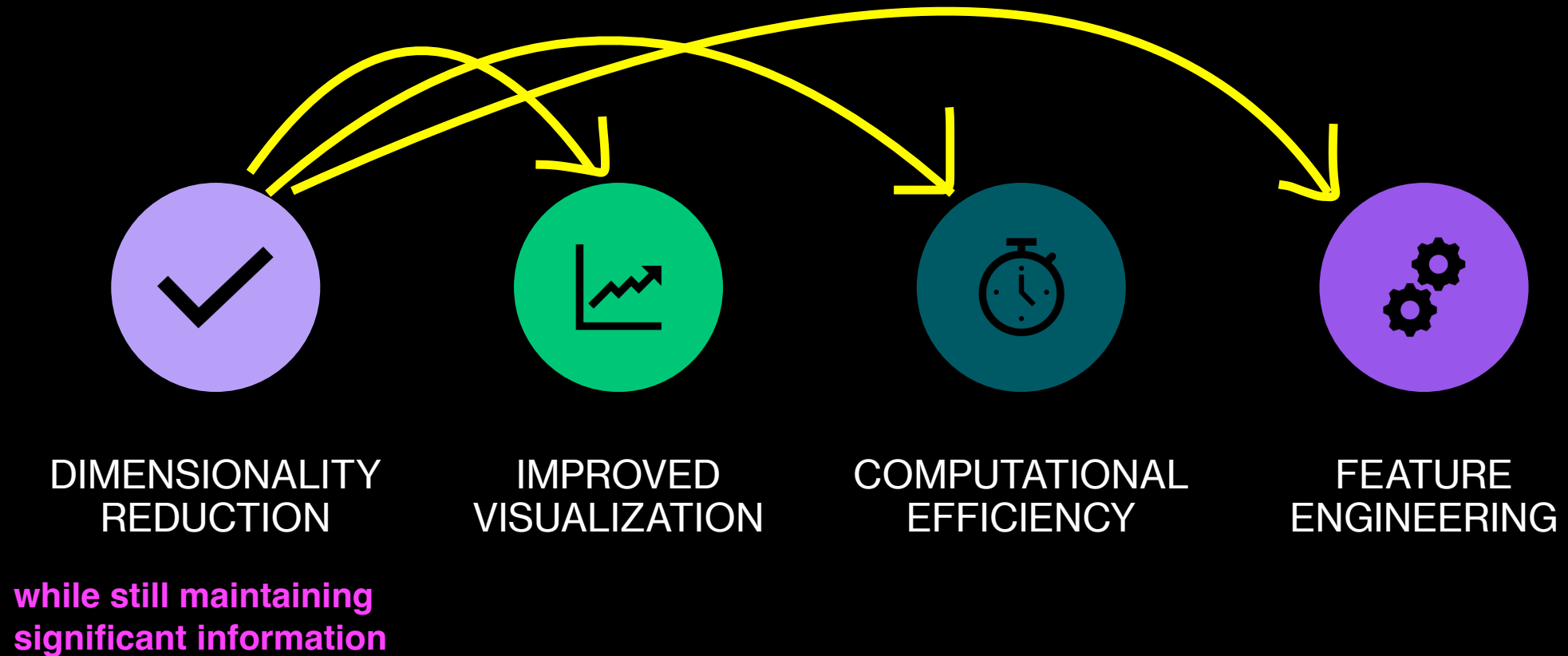
(มี 7 มิติ)

Principal component analysis (PCA) is a dimensionality reduction and machine learning method used to simplify a large data set into a smaller set while still maintaining significant information.

(variance)



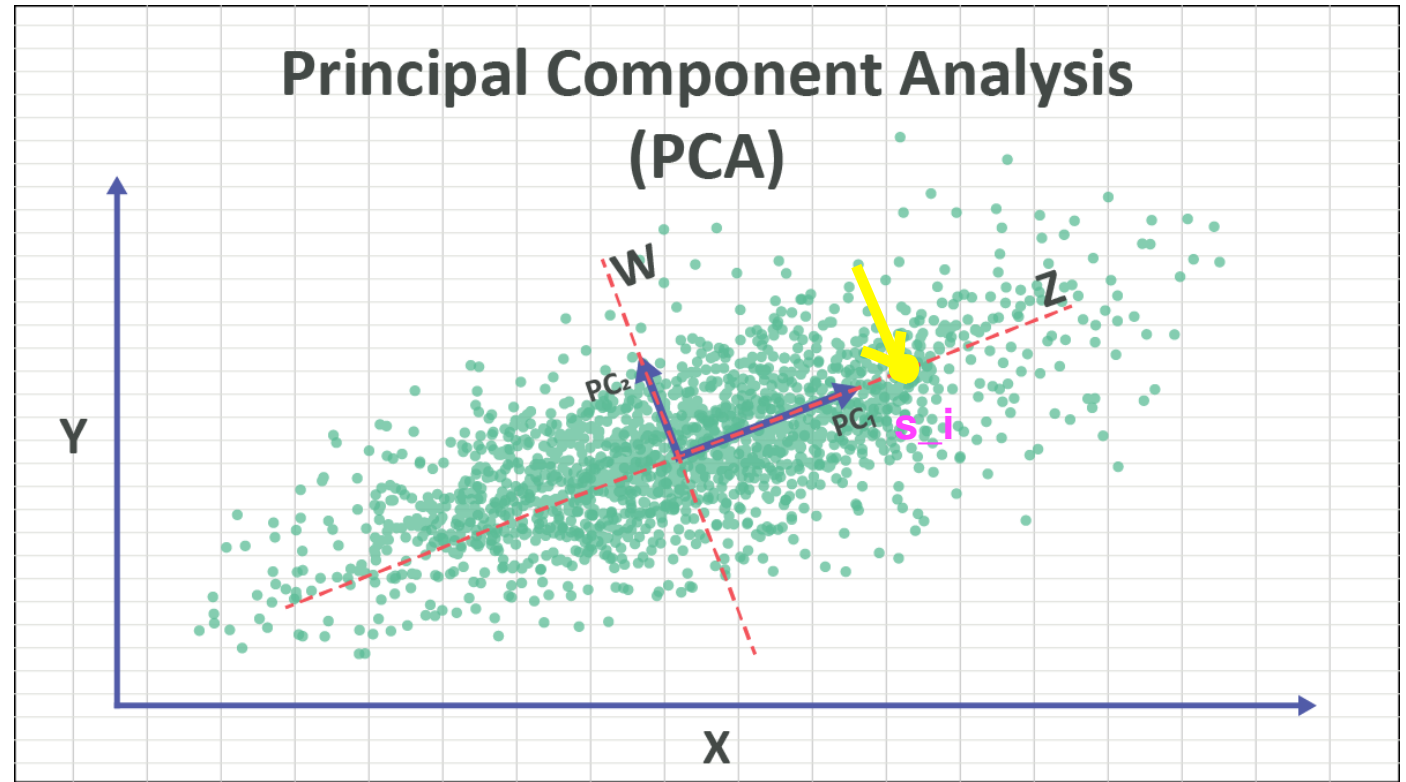
Benefit of PCA



Calculation step

- Standardization
- Compute covariance matrix
- Compute the eigenvalues and eigenvectors
- Consider the eigenvalues **==> dimensionality reduction**
- Map data to new space

Calculation concept



<https://numxl.com/blogs/principal-component-analysis-pca-101/>

ถ้า Matrix ที่คุณอยู่หน้า eigenvector เป็น symmetric matrix
eigenvector ที่ออกมาจะตั้งฉากกันทั้งหมด

เหตุผลที่ทำให้ variance ของระบบก่อนและหลังทำ PCA เท่าเดิม
1. map ไปสู่มิติใหม่ด้วย linear combination
2. eigenvector (v, แกนใหม่) ที่ออกมามันตั้งฉากกันทั้งหมด
3. $\|v\| = 1$

ข้อมูลตั้งต้นมี 5 มิติ \Rightarrow หลังทำ PCA \Rightarrow 5 มิติ
eigenvalues = 20, 7, 6, 5, 2 (40)
= 50%, 17.5%, 15%, 12.5%, 5%
variance บนมิติใหม่จะเป็น
= 50%, 17.5%, 15%, 12.5%, 5%

แกนใหม่!!!

feature ไหนที่มี variance เยอะ \Rightarrow เป็น feature ที่สำคัญ | มาสร้าง feature ใหม่ให้ variance มันเยอะที่สุดกันดีกว่า!!!

Discussion about PCA

มอง variance เป็น information

- What is the motivation of PCA?
- Why PCA axes are orthogonal to each other
- Why summation of variance before PCA and after PCA is still the same?
- Why size of eigenvalue represent size of variance?
- Why Proportion of variance in PCA can be calculated by proportion of eigenvalue?



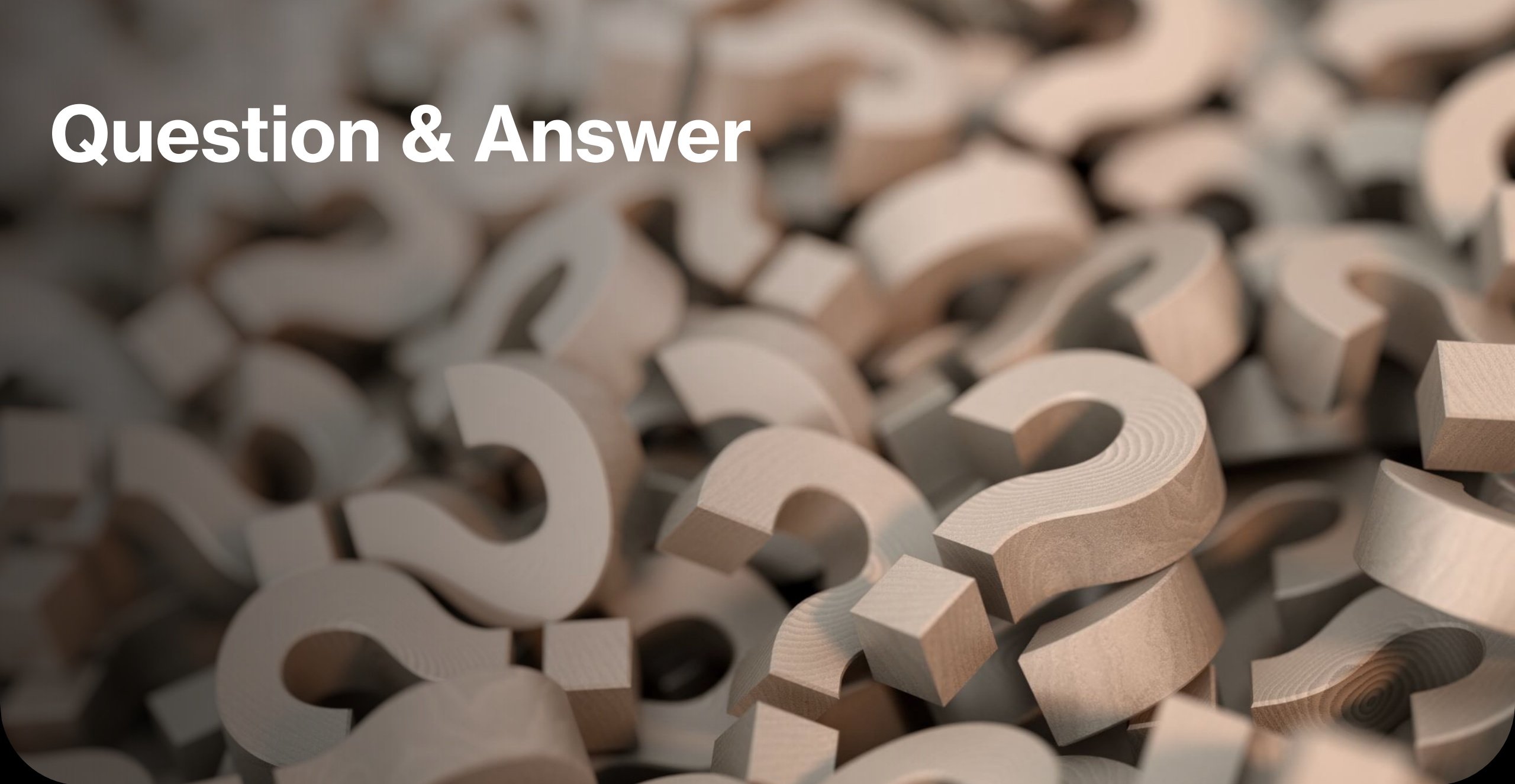
Code: PCA.ipynb



Further reading

- Singular Value Decomposition (SVD)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Linear Discriminant Analysis (LDA)
- Isomap
- Locally Linear Embedding (LLE)

Question & Answer



Reference

- <https://builtin.com/data-science/step-step-explanation-principal-component-analysis>
- <https://numxl.com/blogs/principal-component-analysis-pca-101/>

THANK YOU

