Principal Component Analysis (PCA) for Dimensionality Reduction

Welcome! Today we'll explore how Principal Component Analysis (PCA) helps us reduce data dimensionality without sacrificing valuable information.

By Nisha A K

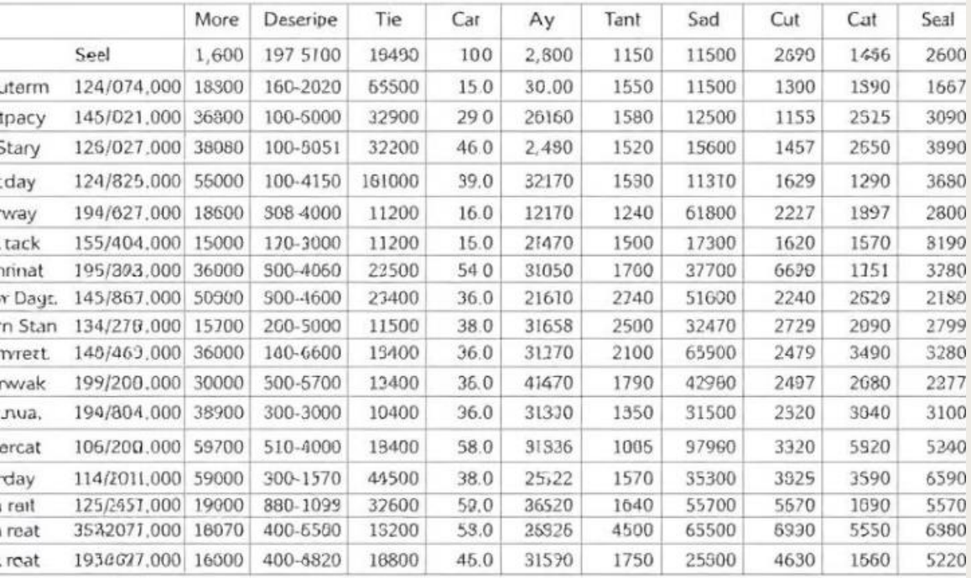
What is Dimensionality Reduction?

Simplifying Complexity

Dimensionality reduction is a technique used to simplify data by reducing the number of variables (dimensions) while preserving meaningful information.

Reducing Noise

It can remove irrelevant or noisy features that might obscure the true underlying patterns in your data.



		More	Deseripe	Tie	Car	Ay	Tant	Sad	Cut	Cat	Seal
	Seel	1,600	197 5100	19490	100	2,800	1150	11500	2690	1456	2600
uterm	124/074,000	18300	160-2020	65500	15.0	30.00	1550	11500	1300	1890	1667
tpacy	145/021,000	36800	100-5000	32900	29 0	26160	1580	12500	1155	2525	3090
Stary	126/027,000	38080	100-5051	32200	46.0	2,490	1520	15600	1457	2650	3990
:day	124/825,000	55000	100-4150	161000	39.0	32170	1530	11310	1629	1290	3680
rway	194/627,000	18600	308-4000	11200	16.0	12170	1240	61800	2227	1897	2800
tack	155/404,000	15000	170-3000	11200	15.0	21470	1500	17300	1620	1570	3190
nrinat	195/393,000	36000	500-4060	22500	54 0	31050	1700	37700	6630	1151	3280
r Dage	145/867,000	50900	500-4600	23400	36.0	21610	2240	51600	2240	2529	2180
rn Stan	134/270,000	15700	200-5000	11500	38.0	31658	2500	32470	2729	2090	2799
nvrett	148/463,000	36000	140-6600	13400	36.0	31270	2100	65900	2479	3490	3280
rwak	199/200,000	30000	500-5700	13400	36.0	41470	1790	42960	2497	2680	2277
.nua.	194/804,000	38900	300-3000	10400	36.0	31330	1350	31500	2320	3040	3100
ercat	106/200,000	59700	510-4000	18400	58.0	31836	1005	97960	3320	5920	5240
rday	114/2011,000	59000	300-1570	44500	38.0	25522	1570	35300	3825	3590	6590
rail	125/2457,000	19000	880-1099	32600	59.0	36520	1640	55700	5670	1090	5570
reat	3542071,000	16070	400-6580	15200	58.0	28926	4500	65500	6930	5550	6380
reat	1936627,000	16000	400-6820	16800	45.0	31590	1750	25500	4630	1560	5220

	Stall	Done	You Last	Eless	Cleyin	Cost	Cidy	Lant	Tabl
	3660							-	
	3568							-	
	9681							-	
	6689							-	
	8090							-	
	5882							-	
	3500							-	
	9572							-	
	3660							-	
								-	
	2400							-	
	2009							-	

Need for Dimensionality Reduction

1 Improved Performance

Reduces computational time for algorithms.

2 Enhanced Visualization

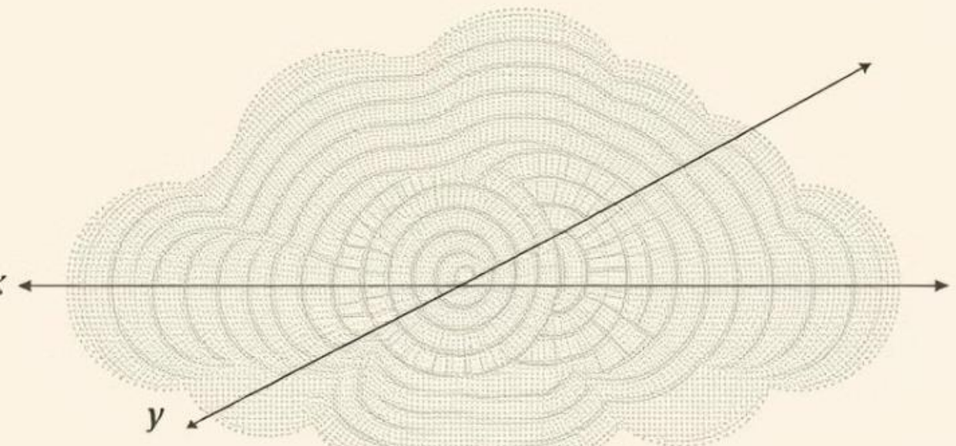
Makes data easier to visualize and interpret.

3 Reduced Storage

Requires less storage space.

4 Avoid Overfitting

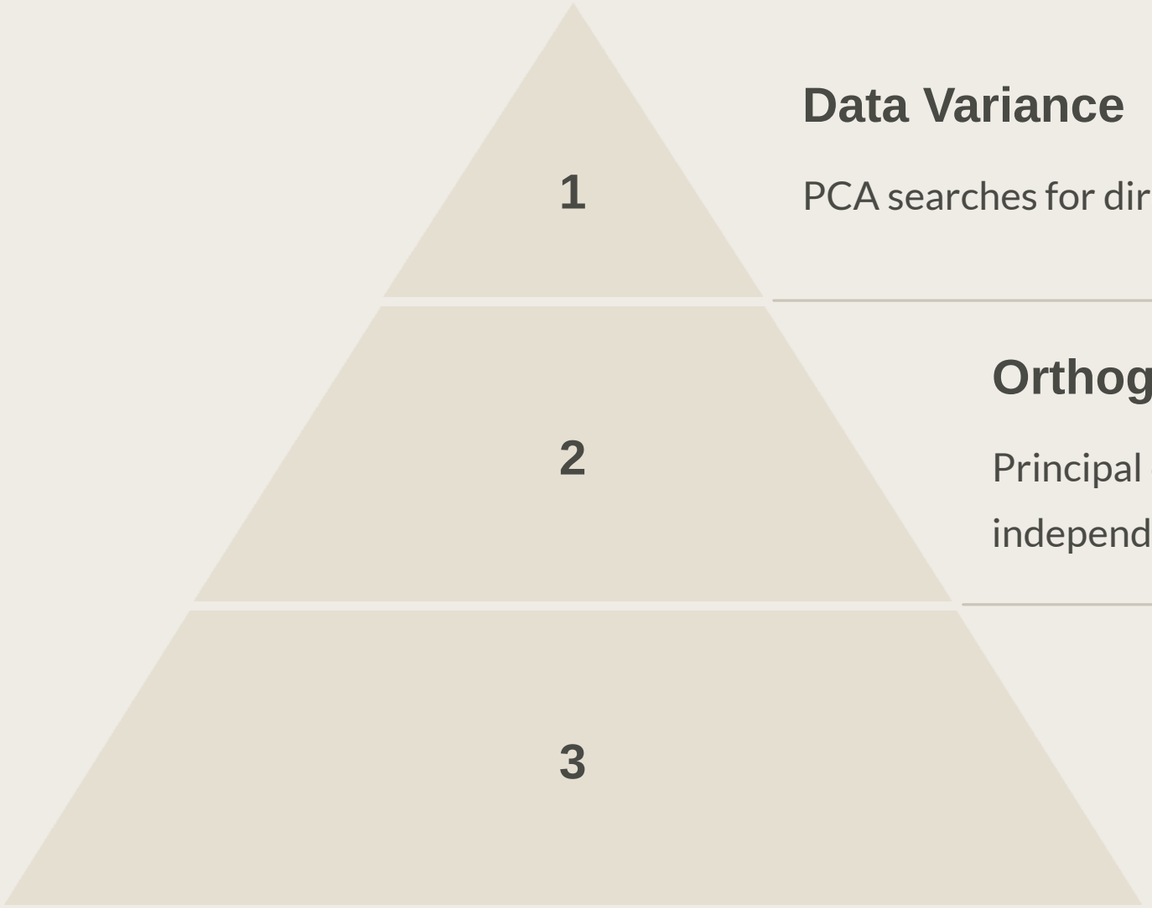
Prevents models from becoming too complex and losing generalization.



Overview of Principal Component Analysis (PCA)

PCA is a powerful technique that finds a new set of orthogonal axes (principal components) that capture the most variance in the data. These components are ordered by their variance, with the first component capturing the most variance.

Geometric Intuition behind PCA



Data Variance

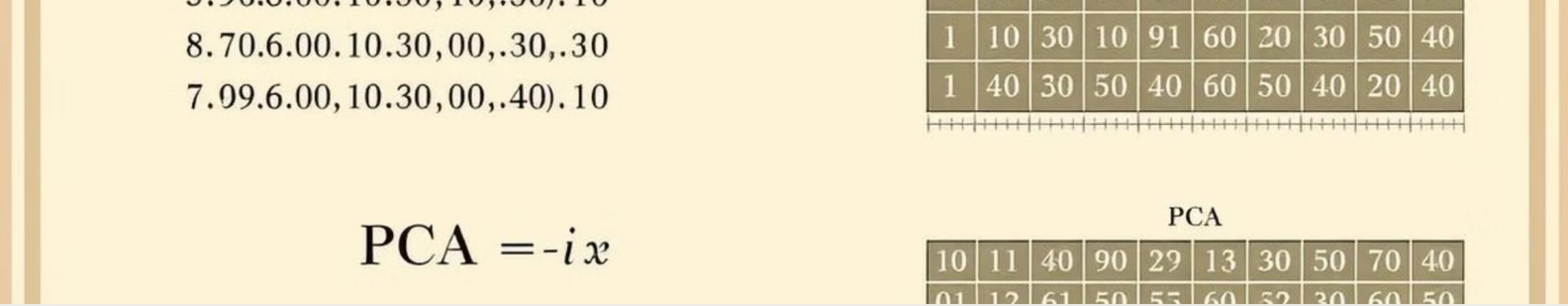
PCA searches for directions of highest data variability.

Orthogonal Axes

Principal components are orthogonal to each other, representing independent directions.

Information Capture

The first few principal components capture most of the data's information.



Calculating Principal Components

PCA involves calculating the covariance matrix of your data and then finding its eigenvectors and eigenvalues. Eigenvectors represent the principal components, and eigenvalues measure the variance explained by each component.

Selecting the Number of Principal Components

You can choose the number of principal components based on the amount of variance you wish to retain. Techniques like the "elbow" method help visualize where diminishing returns set in.

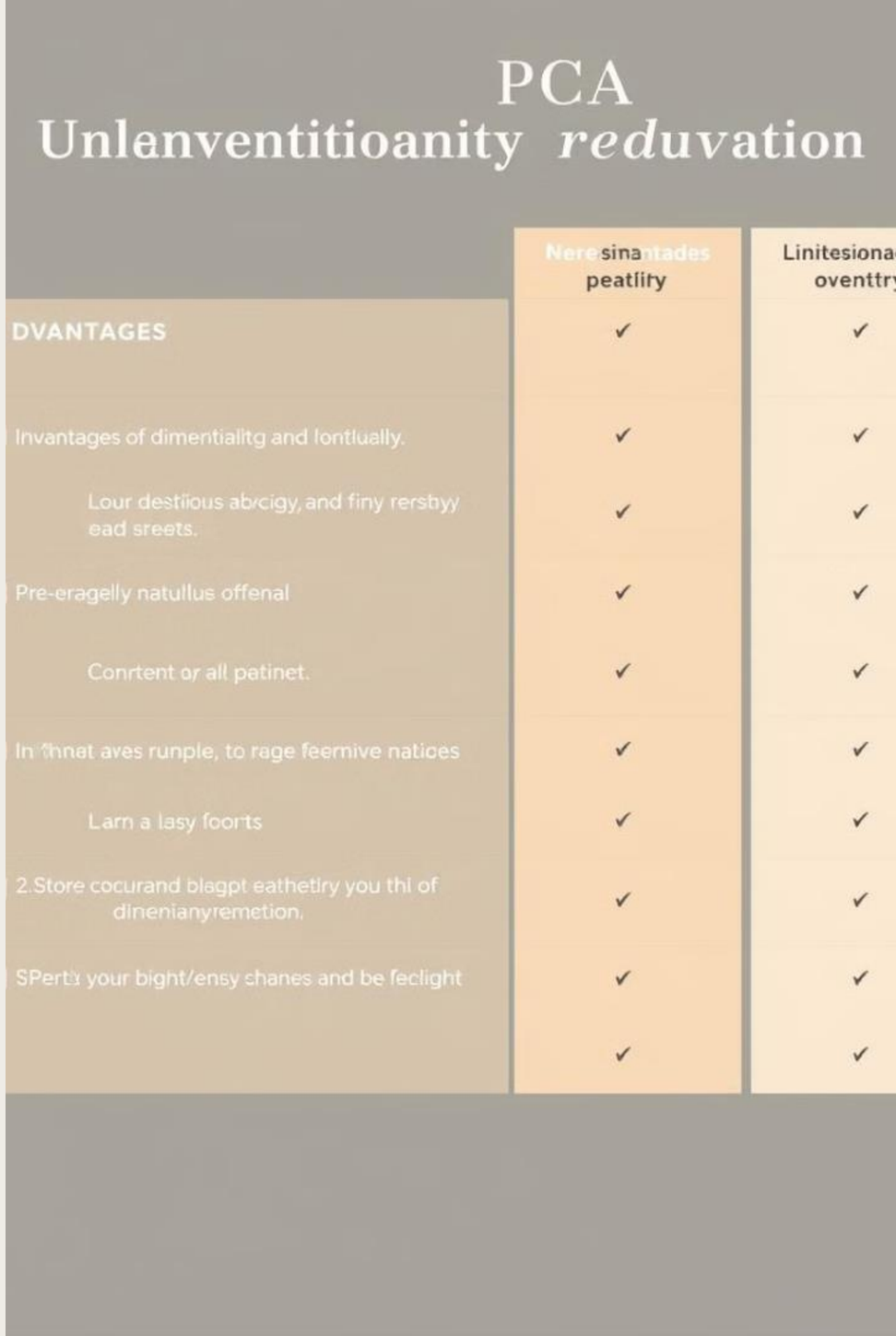
Advantages and Limitations of PCA

Advantages

- Reduces dimensionality
- Improves performance
- Enhances visualization

Limitations

- Loss of interpretability
- Sensitive to outliers
- Assumptions about data distribution



Applications of PCA

Facial Recognition

PCA can be used to identify individuals based on facial features.

Image Compression

PCA can reduce image storage size by removing redundant information.

Financial Analysis

PCA can identify underlying trends in stock market data.

Conclusion and Key Takeaways

PCA is a powerful tool for reducing dimensionality, improving performance, and enhancing understanding of data. Remember to consider its limitations and choose the right number of principal components for your application.