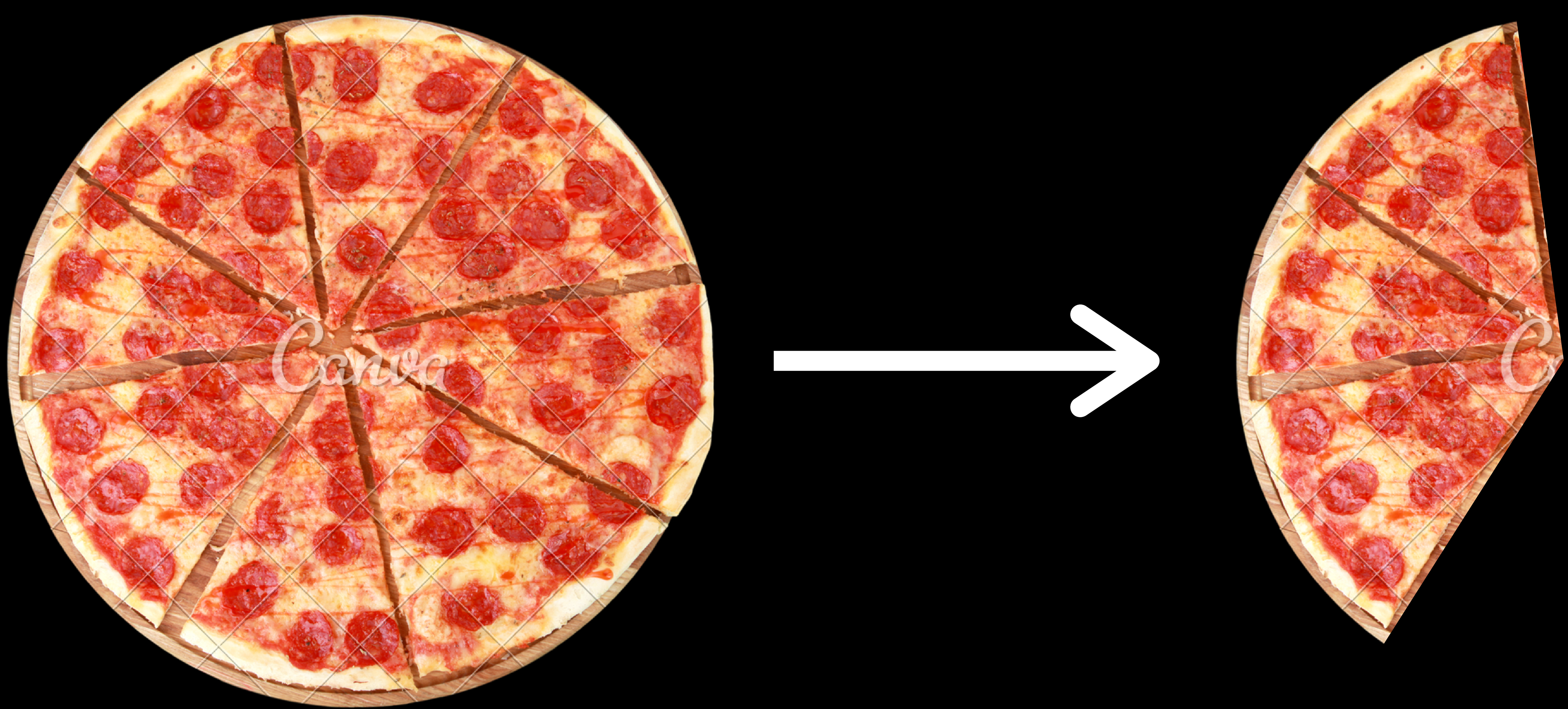
5 mins read

WHAT IS DIMENSIONALITY REDUCTION



What is Dimensionality reduction?

Dimensionality reduction is the process of reducing the number of input variables in training data. Let's say we have N variables in the dataset where we reduce it to K variables (K<<N).

Why Dimensionality reduction?

Few algorithms do not perform well when you have huge amounts of data like KNN, Decision trees etc.. So by reducing them it will help the algorithms to perform well

We cannot visualize data more than 3D so by reducing data to 2D or 3D will allow us to plot and observe patterns more clearly. When we have multicollinearity in our dataset. This technique will remove multicollinearity by removing redundant features.

Components of Dimensionality Reduction

There are two components of dimensionality reduction

Feature selection -: In this, we try to find a subset of the original set of variables, or features, to get a smaller subset which can be used to model the problem. It usually involves three ways:

- Filter
- Wrapper
- Embedded

Feature extraction -: This reduces the data in a high dimensional space to a lower dimension space, i.e. a space with lesser no. of dimensions.

Advantages of Dimensionality Reduction

- It helps in data compression, and hence reduced storage space.
- It reduces computation time.
- It also helps remove redundant features, if any.
- It is helpful in noise removal also and as a result of that, we can improve the performance of models.

Disadvantages of Dimensionality Reduction

- it may lead to some amount of data loss.
- Although, PCA tends to find linear correlations between variables, which is sometimes undesirable.
- Also, PCA fails in cases where mean and covariance are not enough to define datasets.

Different Dimensionality Reduction Techniques?

Linear Dimensionality Reduction Methods

- PCA (Principal Component Analysis)
- Factor Analysis
- LDA (Linear Discriminant Analysis)

Non-linear Dimensionality Reduction Methods

- Multi-dimensional scaling (MDS)
- Isometric Feature Mapping (Isomap)
- Locally Linear Embedding (LLE)
- Hessian Eigenmapping (HLLE)
- Spectral Embedding (Laplacian Eigenmaps)
- t-distributed Stochastic Neighbor Embedding (t-SNE)

Other techniques like

- Autoencoders
- Missing Value Ratio
- Low Variance Filter
- etc....