SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

23AMB201 - MACHINE LEARNING

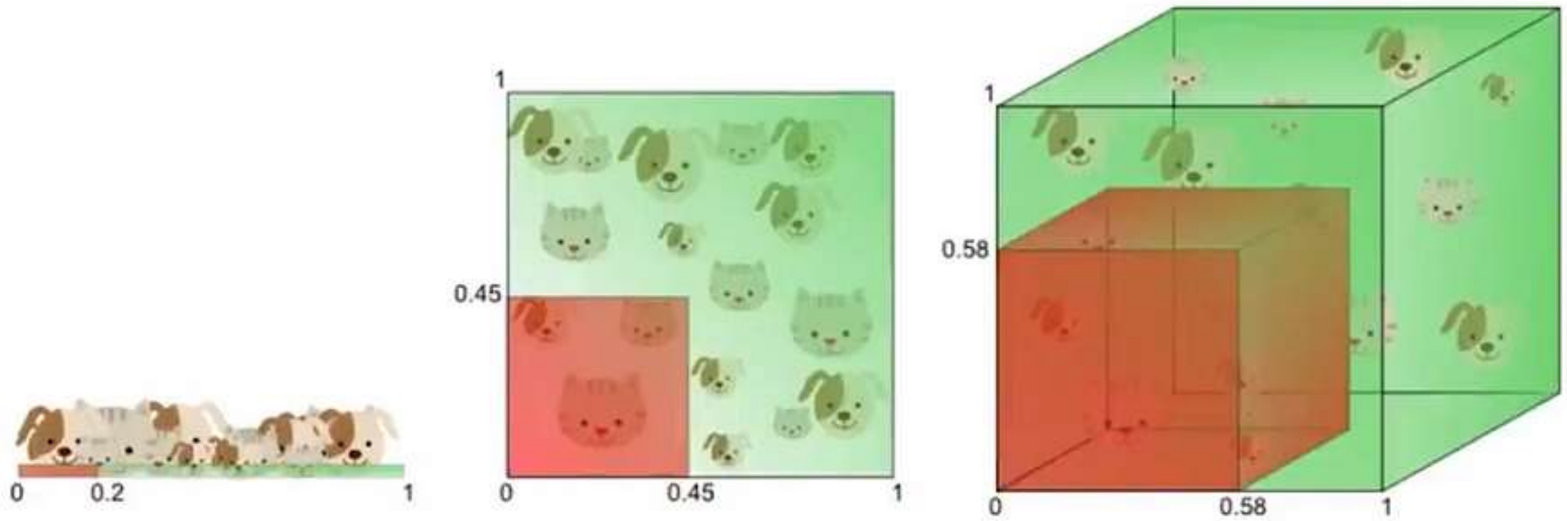
II YEAR IV SEM

UNIT IV – UNSUPERVISED LEARNING ALGORITHM

TOPIC 5 – Dimensionality Reduction

Curse of dimensionality

The "curse of dimensionality" refers to the challenges that arise when analyzing and modeling data with a large number of features (dimensions) in high-dimensional spaces.



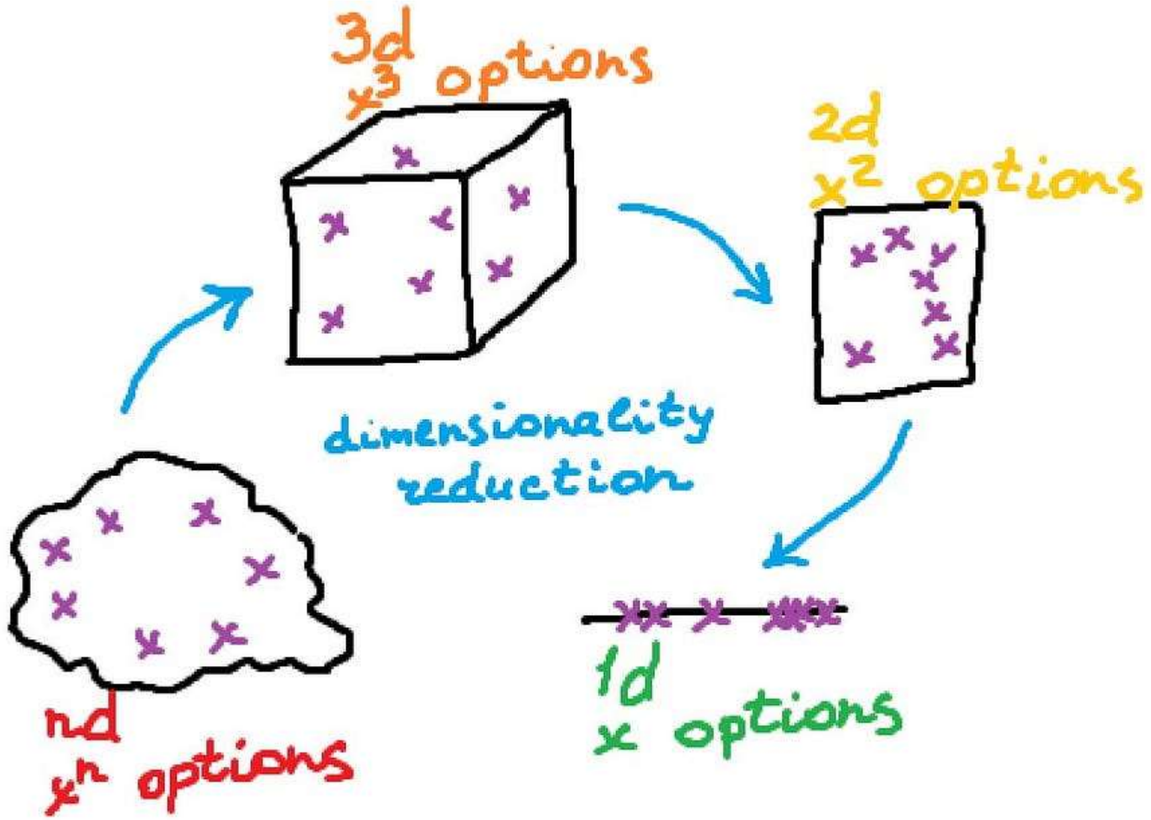
The amount of training data needed to cover 20% of the feature range grows exponentially with the number of dimensions.

Dimensionality Reduction

Dimensionality reduction algorithms represent techniques that reduce the number of features (not samples) in a dataset.

Methods are commonly divided into:

1. Feature Selection — find a subset of the input features
2. Feature Projection (or Feature Extraction) — find the optimal projection of the original data into some low-dimensional space



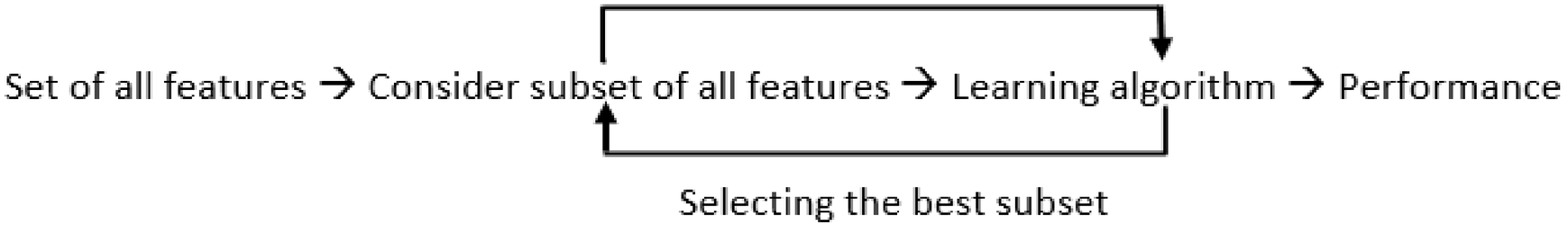
Feature Selection

Techniques used:

1. Filter

Set of all features → Selecting the best subset → Learning algorithm → Performance

2. Wrapper

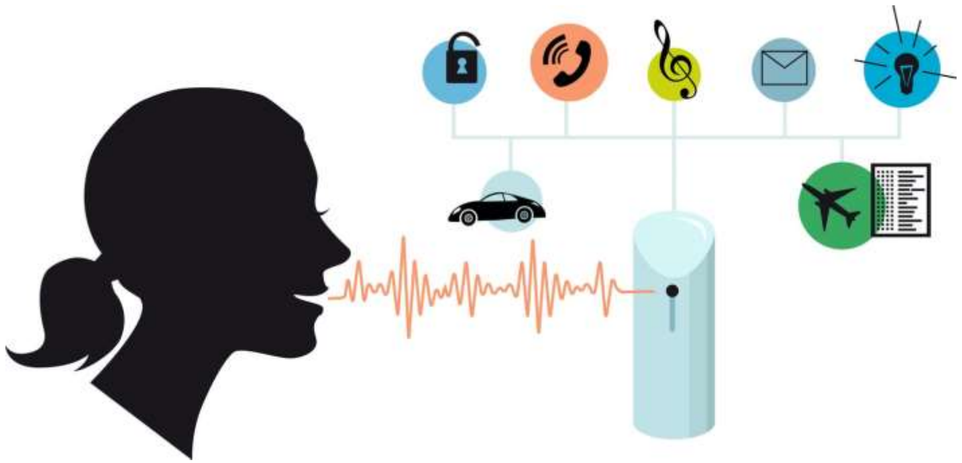


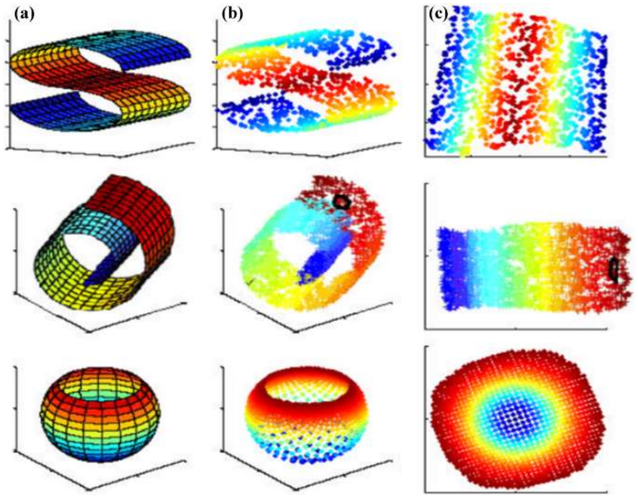
3. Embedded

Applications of DM

Original PNG - 12 MB

Compressed JPEG - 2.5 MB





Techniques

1. Linear Discriminant Analysis
2. Factor Analysis
3. **Principal Components Analysis**
4. Independent Components Analysis

References

Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, ―Learning from Data, AML Book Publishers, 2012.

P. Flach, ―Machine Learning: The art and science of algorithms that make sense of data‖, Cambridge University Press, 2012.

https://www.analyticsvidhya.com/blog/2022/07/principal-component-analysis-beginner-friendly/