Applied Computational Intelligence

Lecture 8 – Cont. Unsupervised Learning

Welcome back

- Pop quiz

This Session

- Unsupervised learning
  - Clustering
    - ✓ K-means
  - Recap Dimensionality Reduction
    - ✓ Feature Extraction
    - ✓ Feature Selection
  - SVD

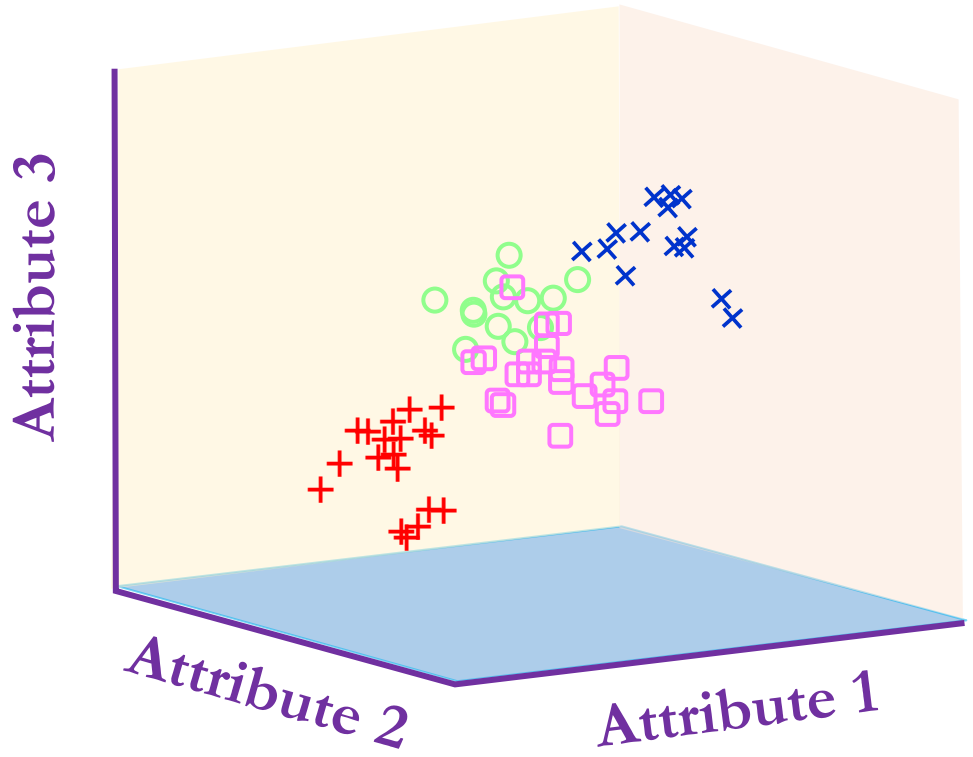
- Next up: supervised learning
  - Linear Regression

Recap - Dimensionality Reduction

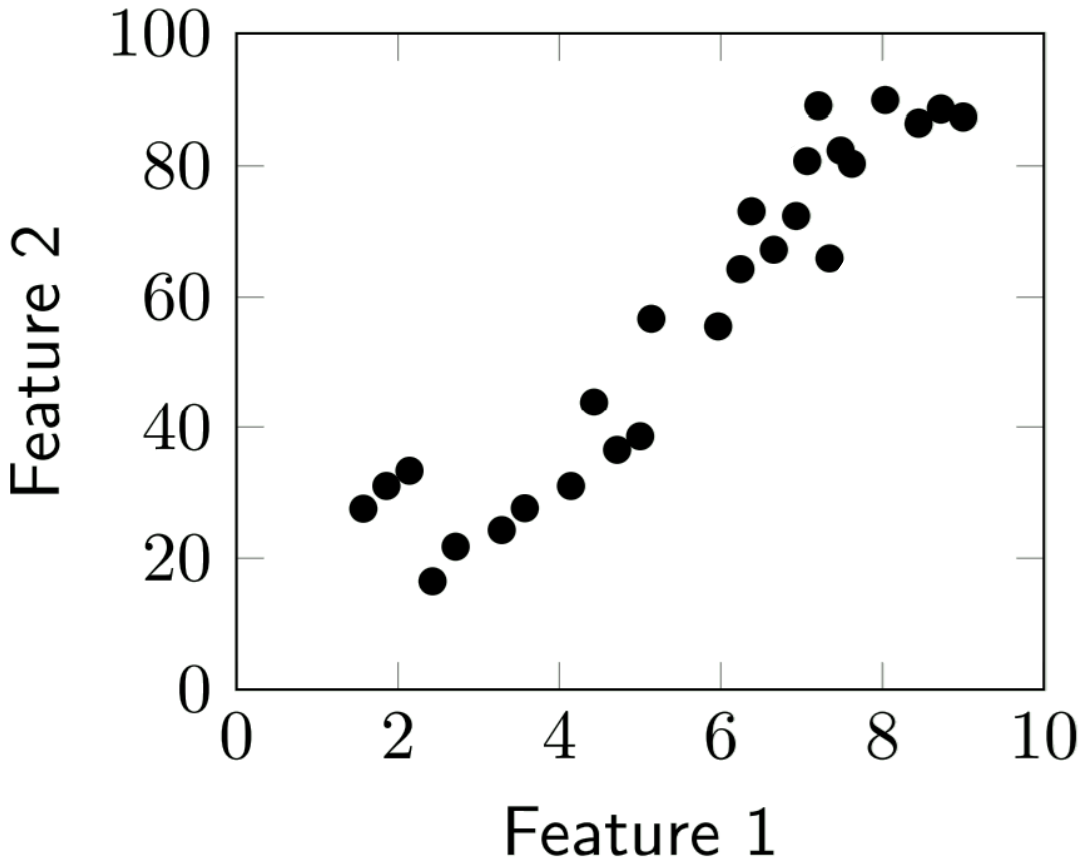
- **Example:**
  - **Input** attributes: a, b, c, d
  - **Target** attribute: t
    - ✓Model: $m(\{a, b, c, d\}, \{t\})$
  - **Attribute construction**
    - ✓New attribute: $z = f(a, b)$
    - ✓Model: $m(\{z, c, d\}, \{t\})$
  - **Attribute subset selection**
    - ✓If we find c has **insignificant discriminability** between clusters/ classes or c ≈ 0. So, it can be removed the model
    - ✓$M(\{a, b, d\}, \{t\})$

Dimensionality Reduction Cont.

3D Attribute Space

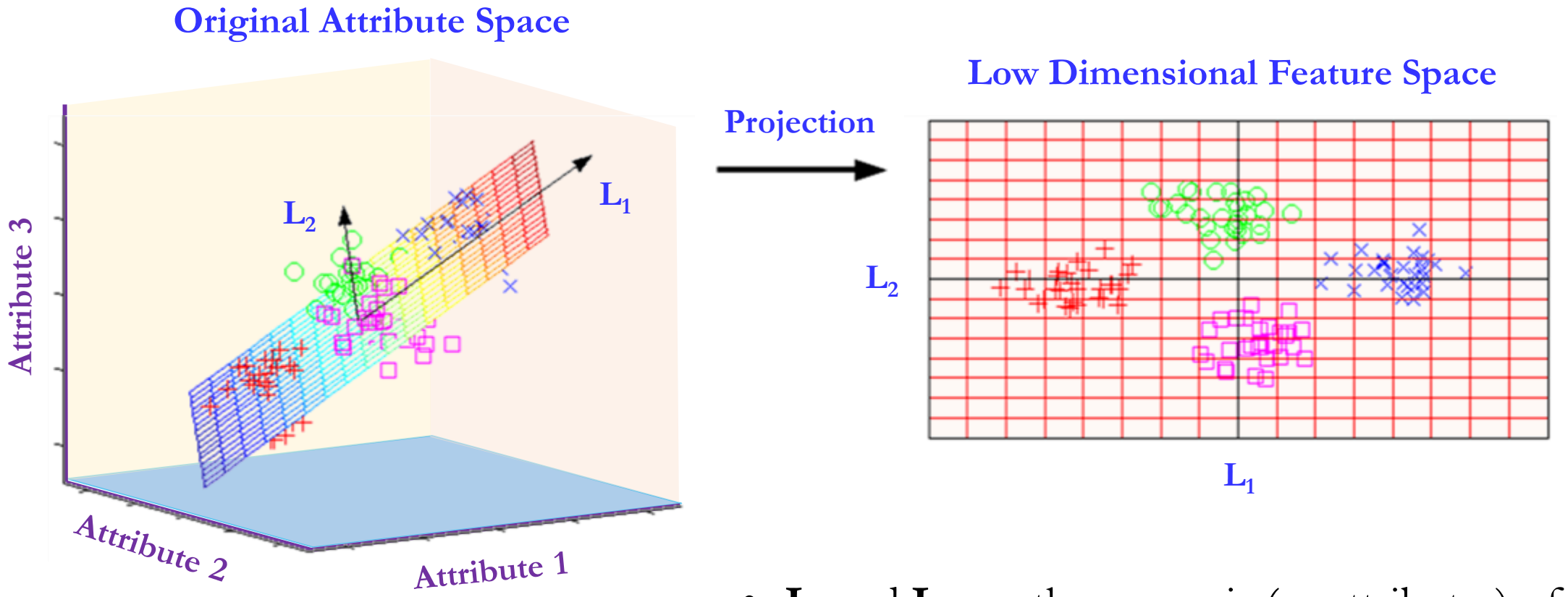


2D Attribute Space



- Can we **find a low dimensional subspace** where **all** the high dimensional **data points be embedded**?

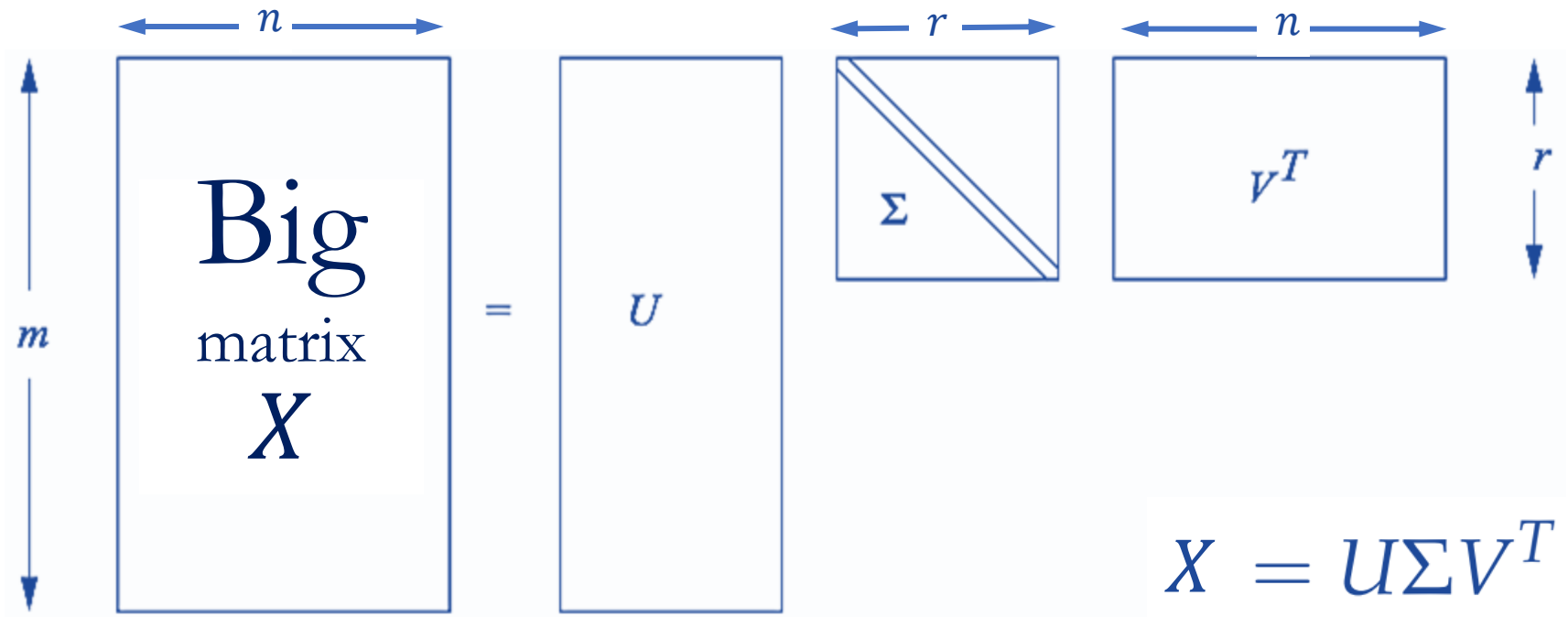
Dimensionality Reduction Cont.



- $\mathbf{L_1}$ and $\mathbf{L_2}$ are the new axis (or attributes) of the constructed low dimensional attribute space.

Dimensionality Reduction Using SVD

- It is a mathematical procedure that performs a matrix decomposition (factorization)
- It represents a **big matrix X into product of 3 small matrices**: U, Σ, and V^T



Dimensionality Reduction Using SVD - Example

$$X \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = U \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \Sigma \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times V \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

- **New representation of the data**
  - Drop the vectors in U and V associated with the smallest singular value

Dimensionality Reduction – Summary

- **Discover** hidden **topics** or **correlations**
  - E.g., words that occur commonly together
- **Remove redundant** attributes and noisy (outlier) attributes
  - Not all words have inherent information
- **Interpretation** and visualization
- **Reduced complexity**
  - Easier storage and data processing

Slide adopted from Jure Leskovec, CS246: Mining Massive Datasets, Stanford University

Applications of Unsupervised Learning

- **News and article organization: E.g.,** Google News uses unsupervised learning to categorize articles on the same story from various online news outlets.
- **Computer vision:** E.g., visual perception tasks, such as object recognition.
- **Healthcare:** E.g., diagnosing patients quickly
- **Anomaly detection:** E.g., raising awareness around faulty equipment, human error, or security breaches
- **Customer behavior analysis and recommendation systems:** E.g., understanding common traits and purchasing habits.

Pop Quiz

- Time for pop quiz

Now…

- Discussion on Assignment

References

1. Eberhart, Russell C., and Yuhui Shi. Computational Intelligence : Concepts to Implementations, Elsevier Science & Technology, 2011.
2. Stuart J. Russell and Peter Norvig, Artificial Intelligence: A Modern Approach,4th Edition, 2020.
3. Teknomo Kardi, "K-Means Clustering: Numerical Example". In: (2017). url: http://people.revoledu.com.