Unsupervised and Unstructured Machine Learning
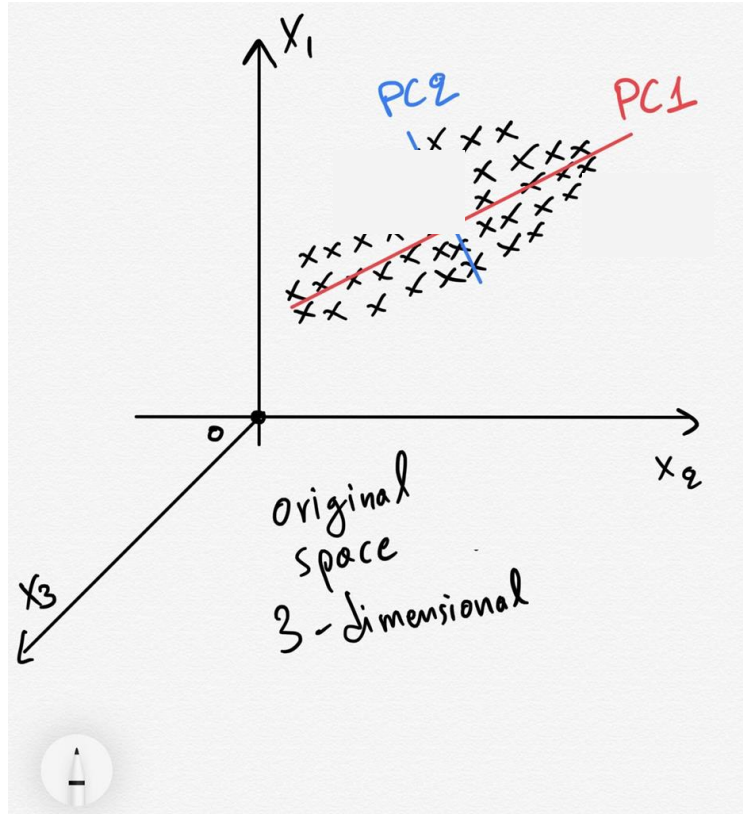
BA820 – Mohannad Elhamod

Feature Reduction

Problems with having too many attributes:

- Analyses/Modeling can take a very long time

- Data can take too much space.

- Risk of correlation/redundancy amongst the variables.
  - Difficulty in interpreting the fit of our models.
  - Tend to overemphasize the underlying variable's contribution.

- Helps remove noise.

- Not easy to visualize/interpret.

- Curse of Dimensionality!

Boston University Questrom School of Business

Dimensionality Reduction

**Boston University** Questrom School of Business

Dimensionality Reduction Techniques

- **Linear:**

  - The new dimensions are _a linear_ combination of the originals.

  - PCA is the prime example of this category.

- **Non-linear:** such, as t-SNE and UMAP.

Principal Component Analysis

Intuition

- Goal is to find direction(s) that captures the largest amount of variation in data.
- We call these direction(s) *principal component(s).*

Boston University Questrom School of Business
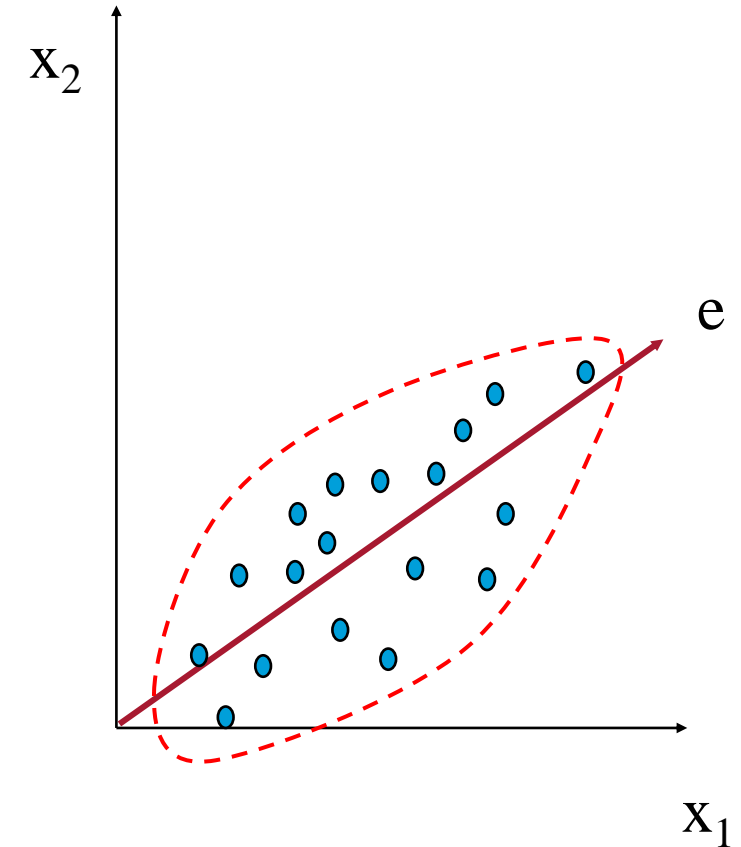
Properties of Principal Components

- First component lies along the direction of the data's largest variance/spread.

- Each component is perpendicular to all other components (independence).

- The components are ordered in terms of their ability of explaining the data (i.e., in order of how much variance in the data they capture).

- Let's play with this

Boston University Questrom School of Business

Dimensionality Reduction with PCA

- The number of components available is equal to the number of attributes being analyzed.

- However, in *most* analyses, only the first few components account for meaningful amounts of variance (>90%), so only these first few components are retained, interpreted, and used in subsequent analyses.

- When you remove dimensions. You lose some information!

How Many Components do we select?

- <u>Method1:</u> Remove dimensions when reaching a sufficient cumulative explained variance.

- <u>Method2:</u> Elbow method could be used.

Scree plot
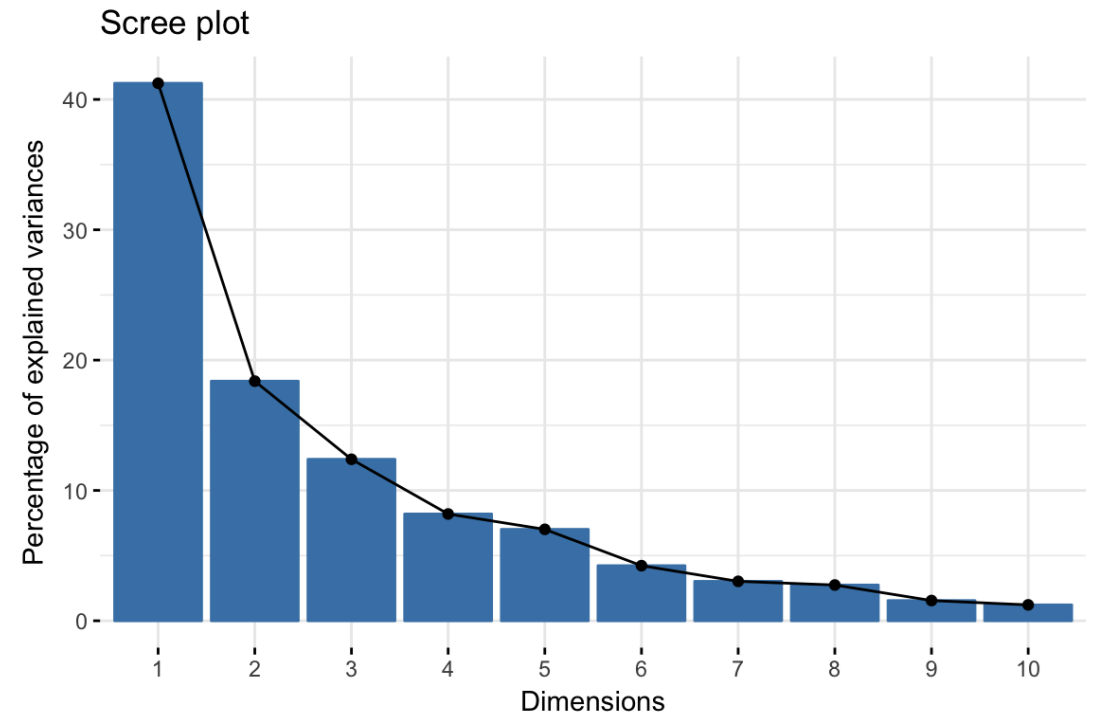
Considerations

- PCA assumes the data has linear patterns in terms of the original attribute.

- PCA can be used for Feature Engineering (the new features can be used for down-stream tasks).
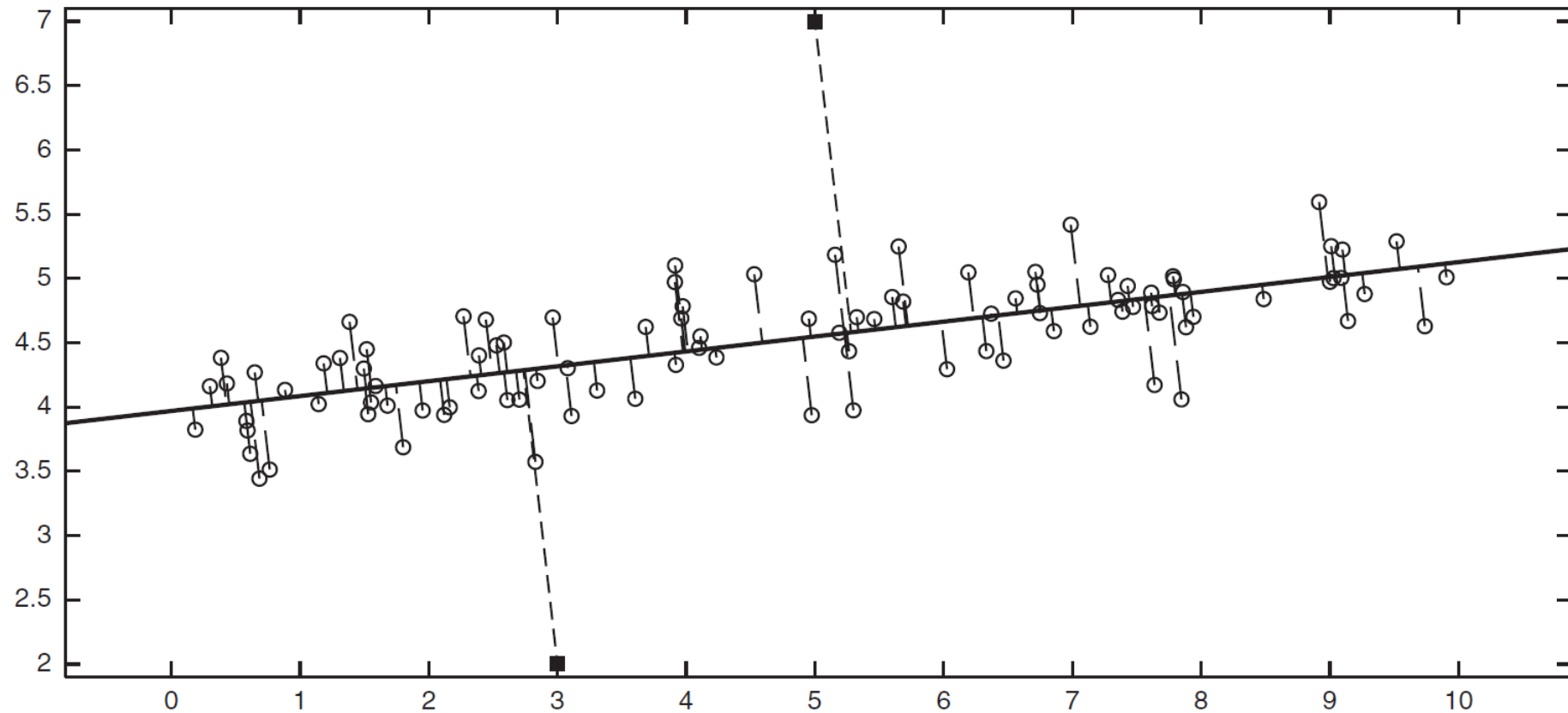
Boston University Questrom School of Business

Reconstruction Error

- Let **x** be the original data point.
- Using PCA, project the point to a lower dimensionality.
- Project the object back to the original space. Call this object $\hat{\mathbf{x}}$

$$\text{Reconstruction Error(x)} = \|x - \hat{x}\|$$

- Points with large reconstruction errors are anomalies

BOSTON
UNIVERSITY

Reconstruction of two-dimensional data

Boston University Questrom School of Business

Demo

Boston University Questrom School of Business