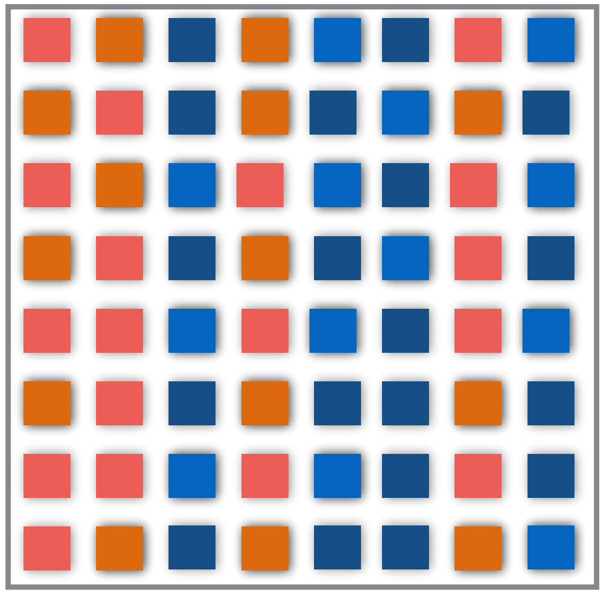
Fusion Subspace Clustering: Full & Missing Data

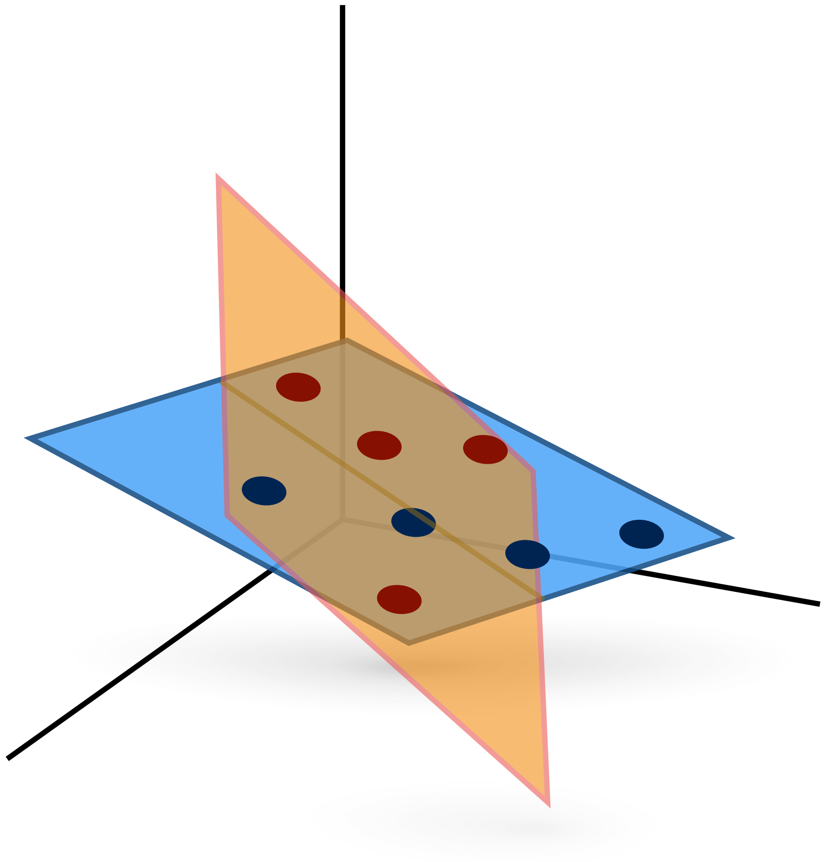
Daniel Pimentel-Alarcón
Computer Science
Georgia State University

Fusion Subspace Clustering: Full & Missing Data

Daniel Pimentel-Alarcón
Computer Science
Georgia State University

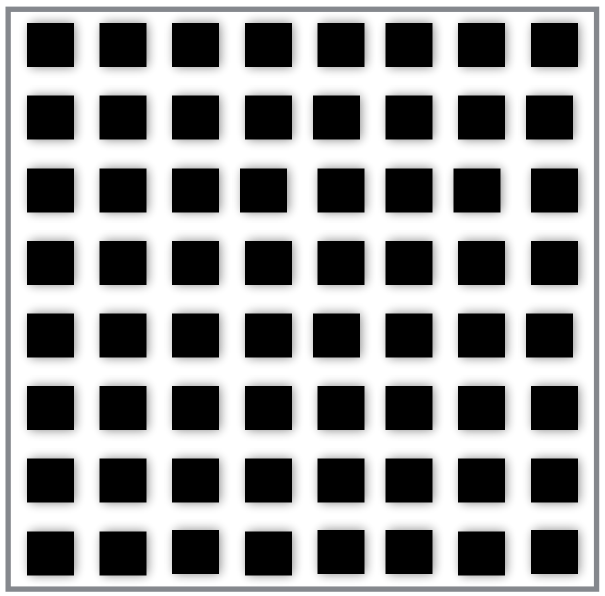
Joint work:
Usman Mahmood

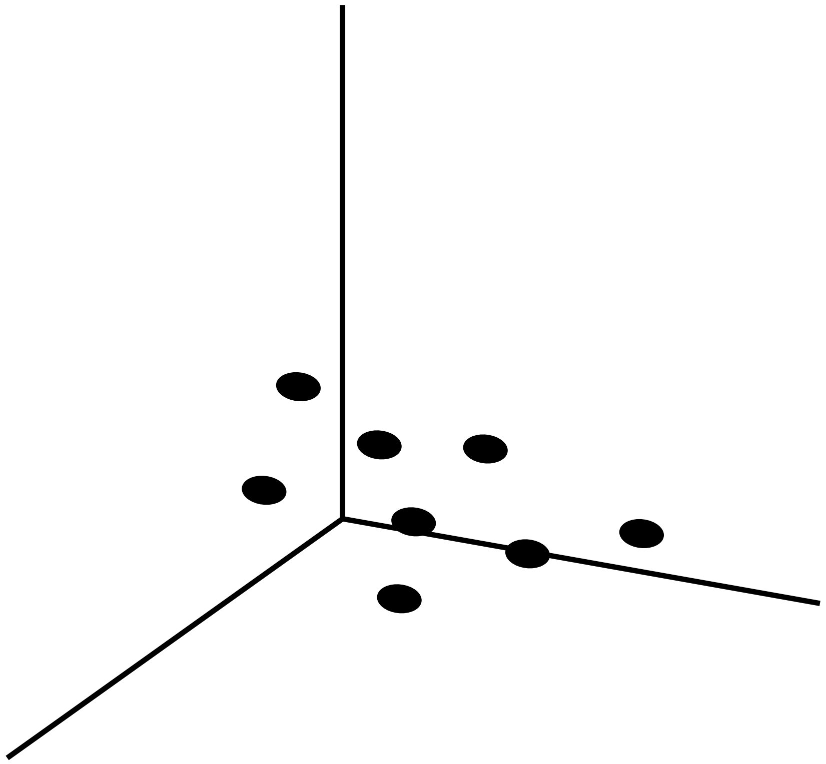




Subspace Clustering

Goal: Cluster columns

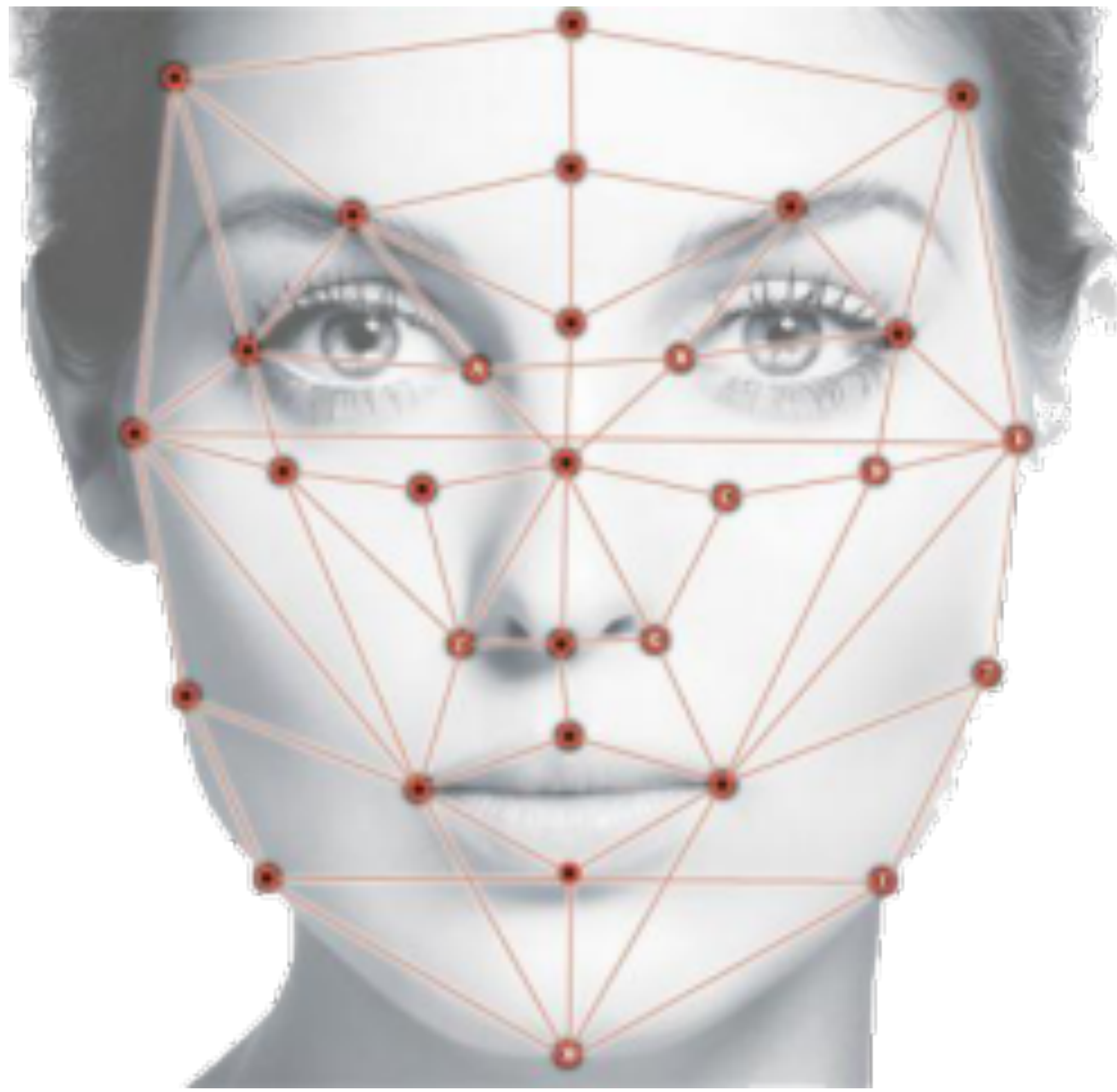




Subspace Clustering

Goal: Cluster columns


What is this good for?



Lots of Applications

Face Clustering

Lots of Applications

Motion Segmentation

Fusion Subspace Clustering

This talk

THE FOLLOWING **PREVIEW** HAS BEEN APPROVED FOR

ALL AUDIENCES

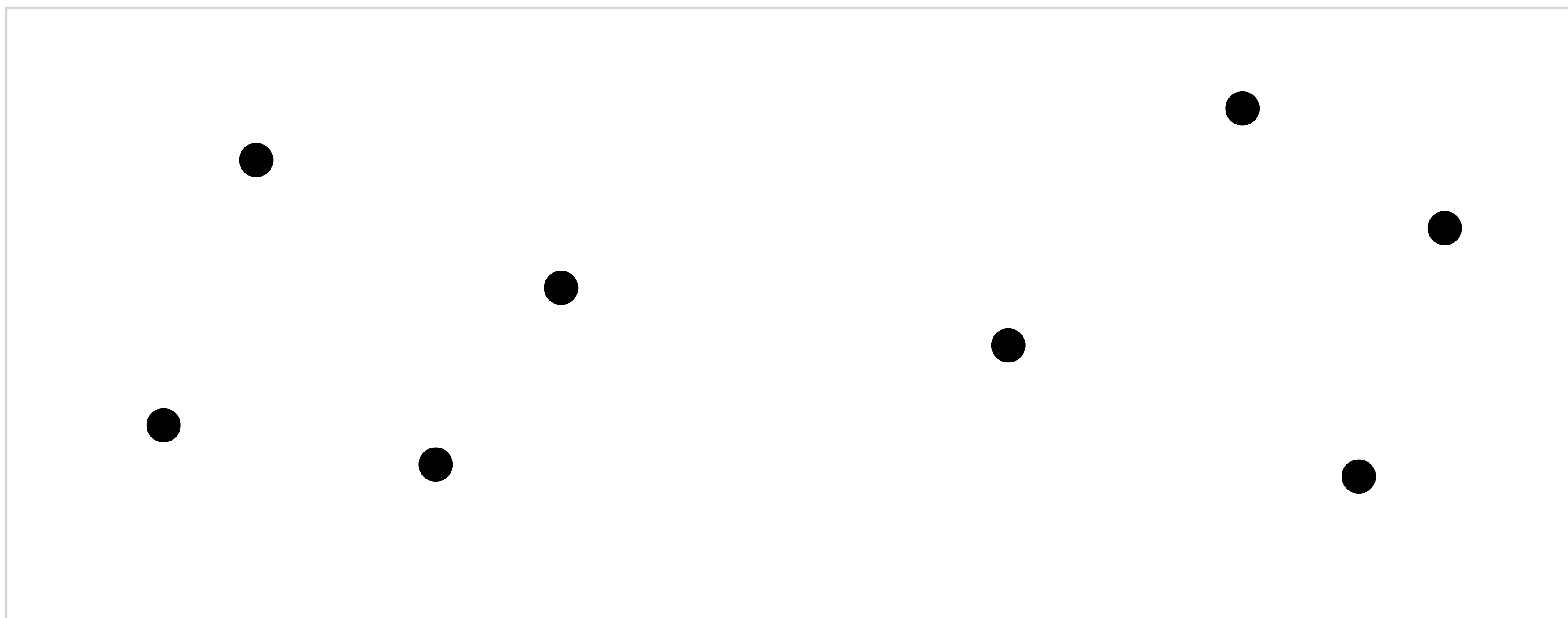
BY THE MOTION PICTURE ASSOCIATION OF AMERICA INC.

THE FILM ADVERTISED HAS BEEN RATED

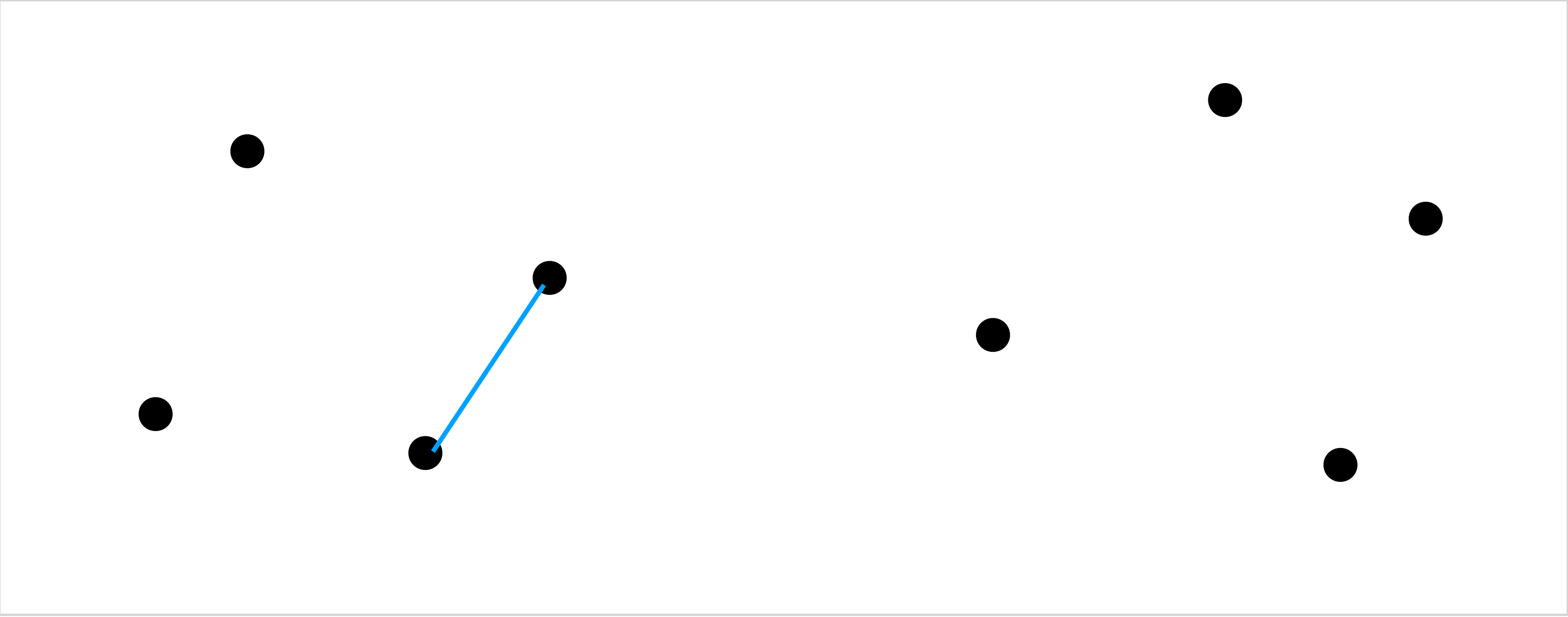
R	RESTRICTED
	UNDER 17 REQUIRES ACCOMPANYING PARENT OR GUARDIAN
PARTIAL NUDITY & OPTIMIZATION	

www.filmratings.com

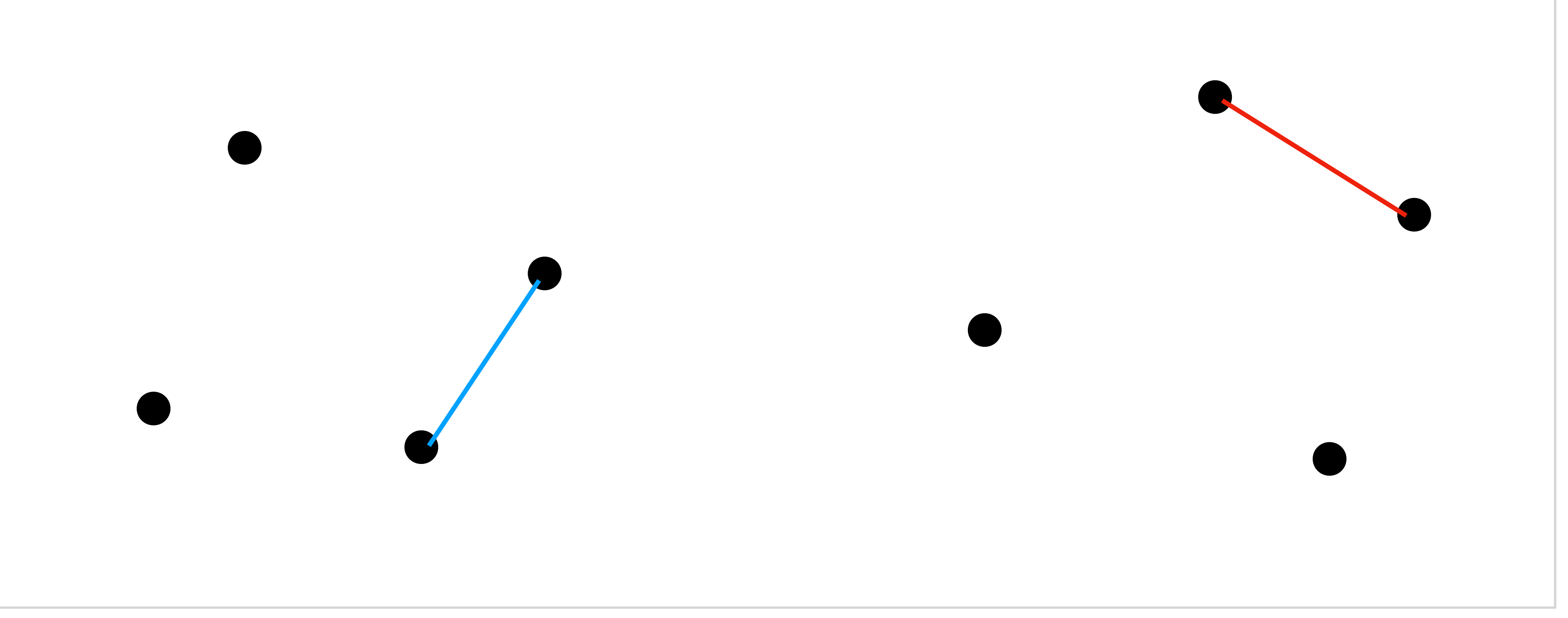
www.mpaa.org



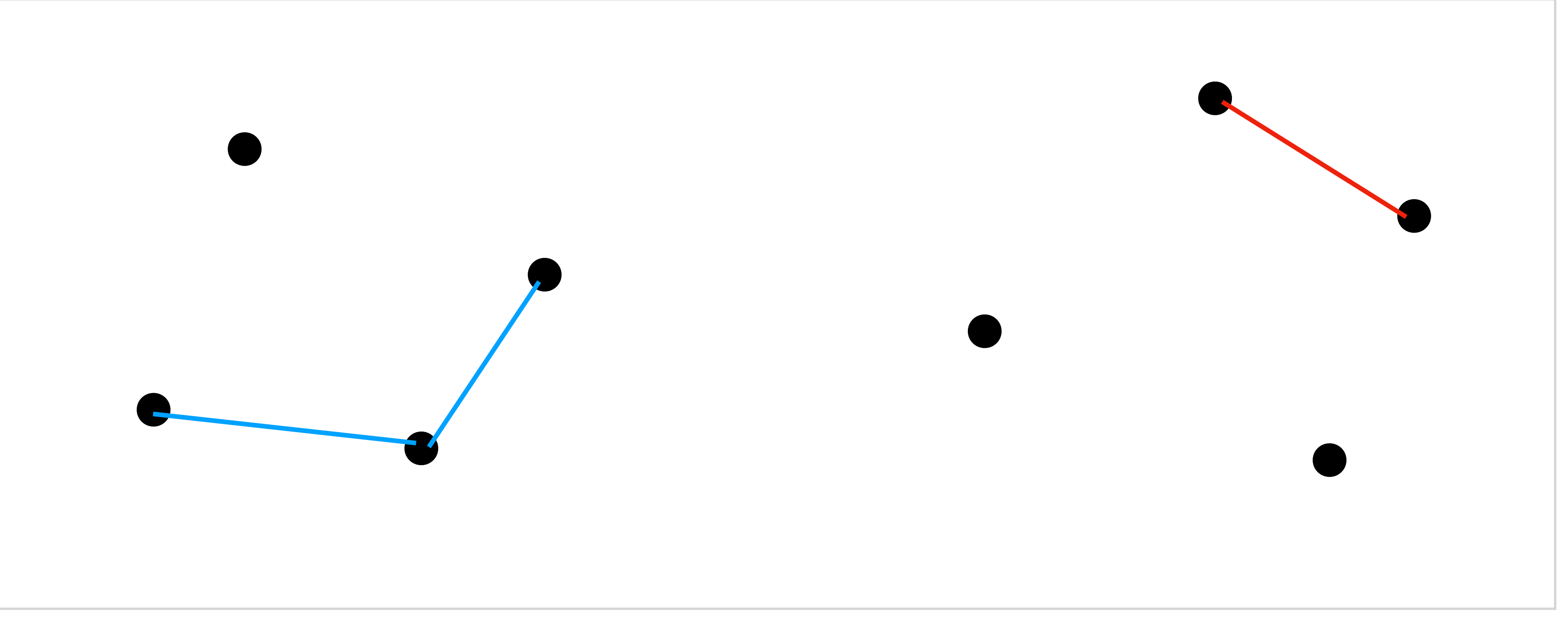
Hierarchical Clustering



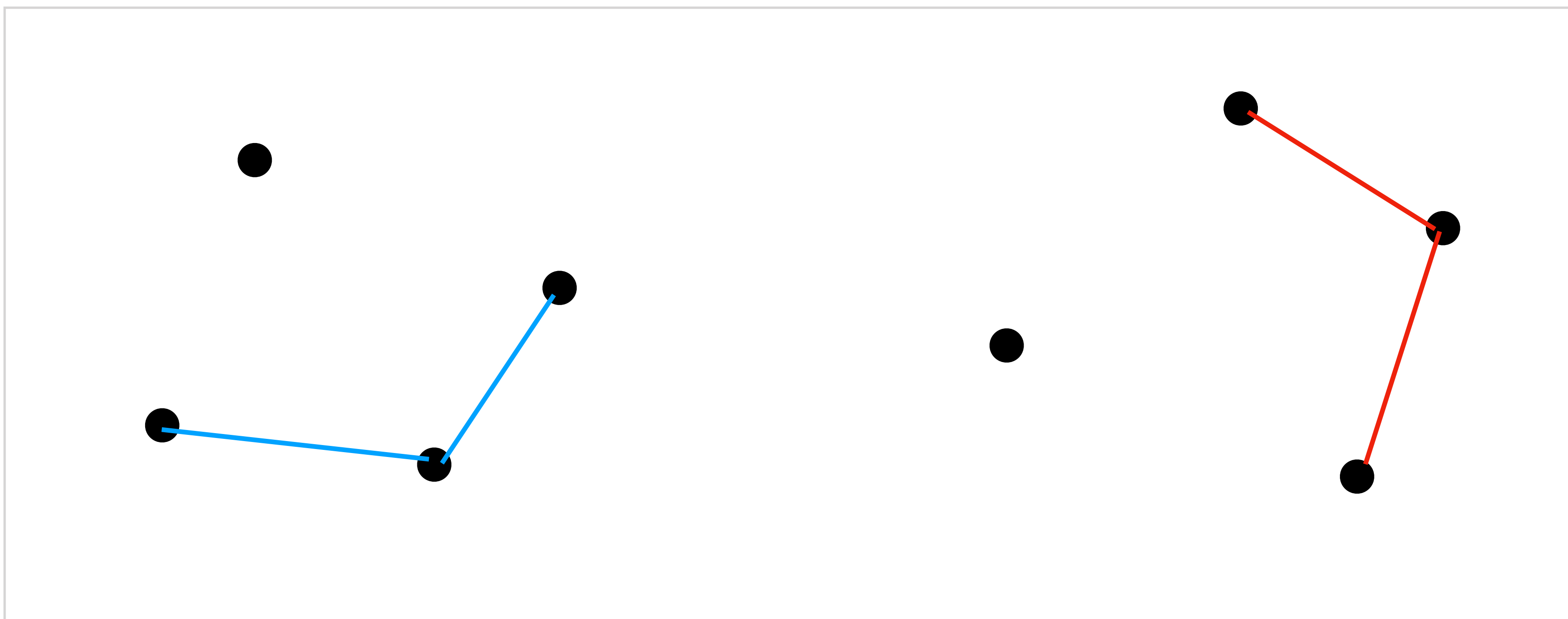
Hierarchical Clustering



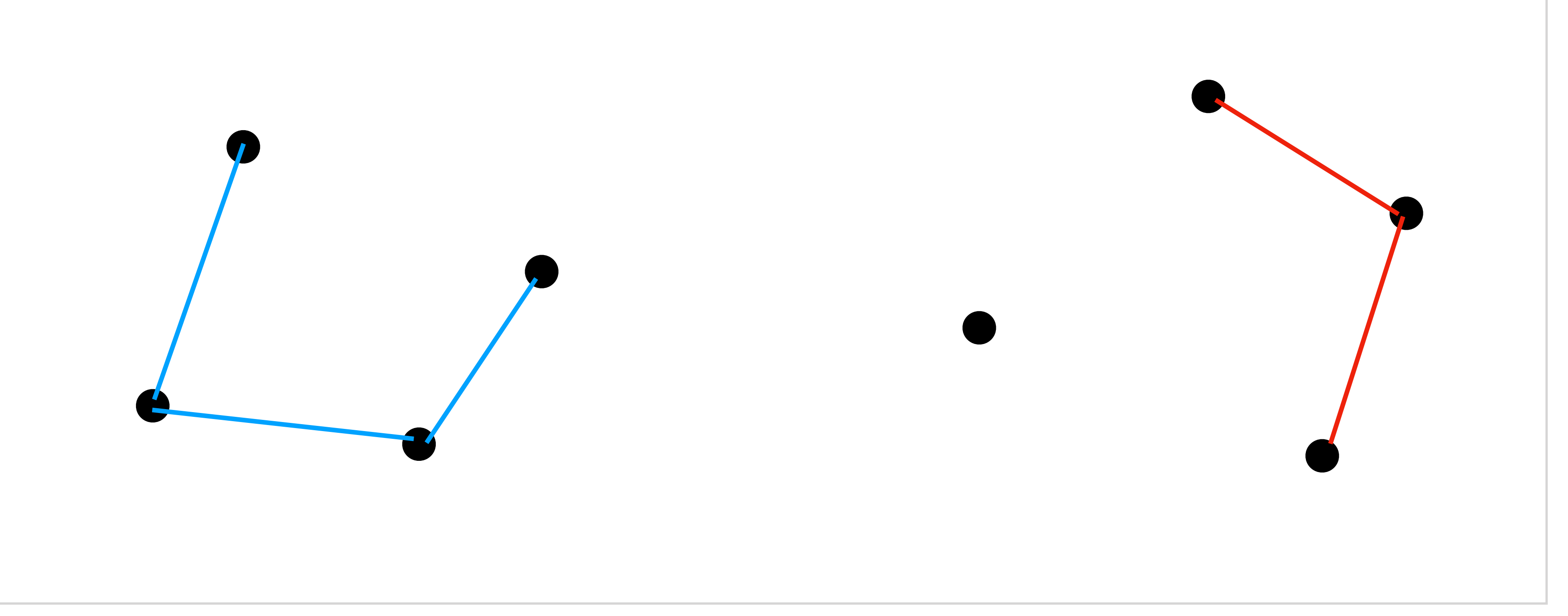
Hierarchical Clustering



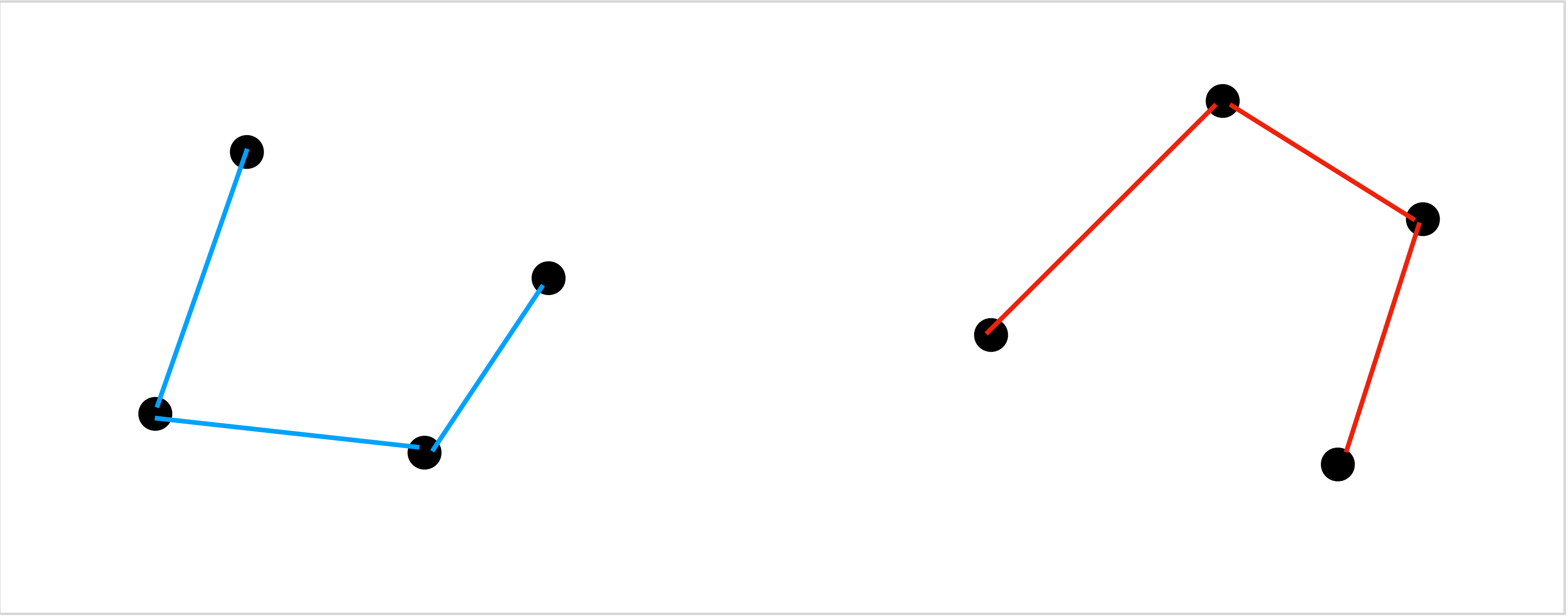
Hierarchical Clustering



Hierarchical Clustering



Hierarchical Clustering

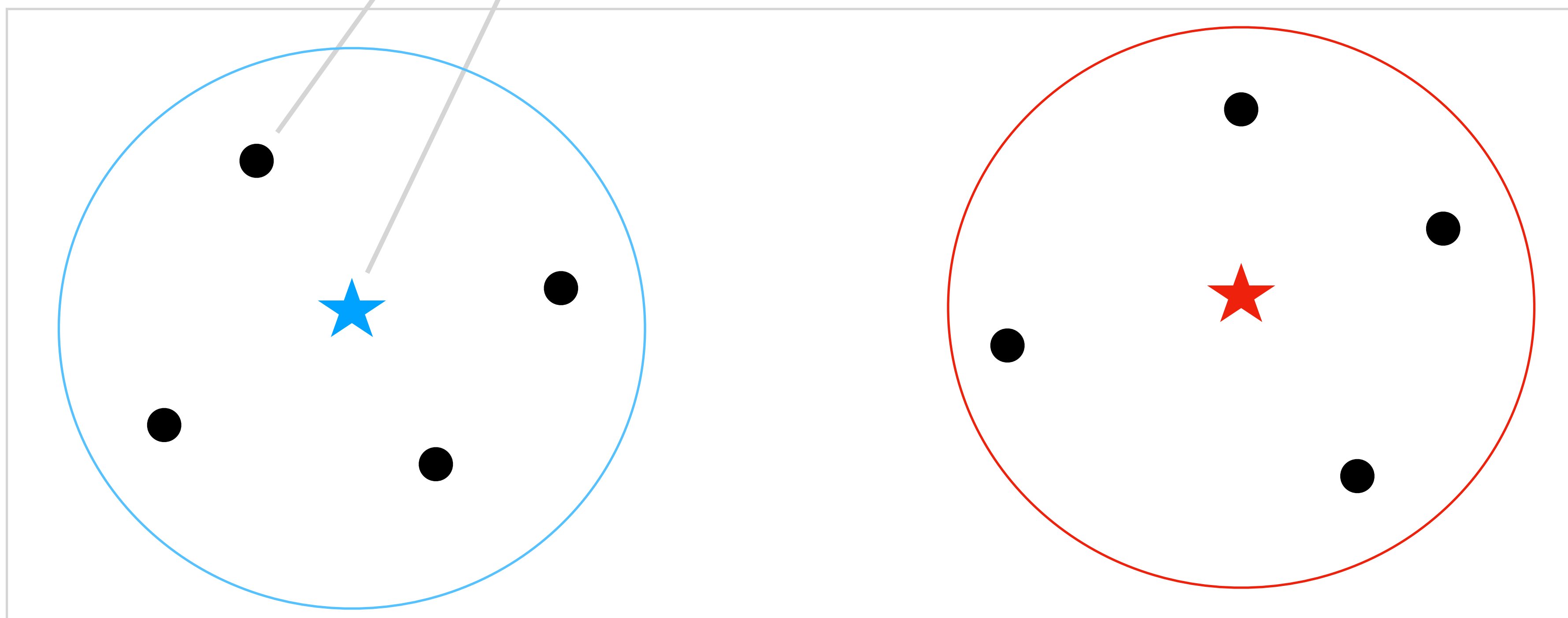


Hierarchical Clustering

Number of points

Number of Clusters

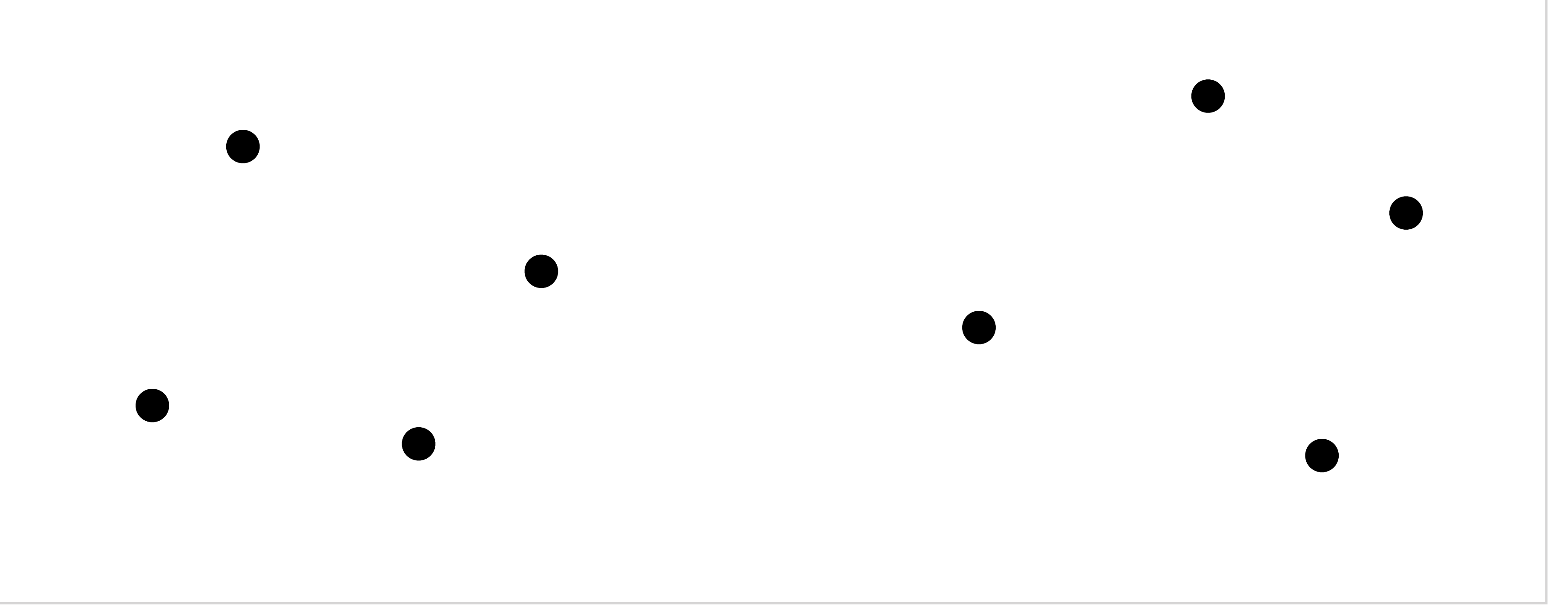
$$\underset{\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{y}_i - \boldsymbol{\mu}_i\|_2^2 \quad \text{subject to} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{1}_{\{\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j\}} \leq K$$



Hierarchical Clustering

As an Optimization Problem

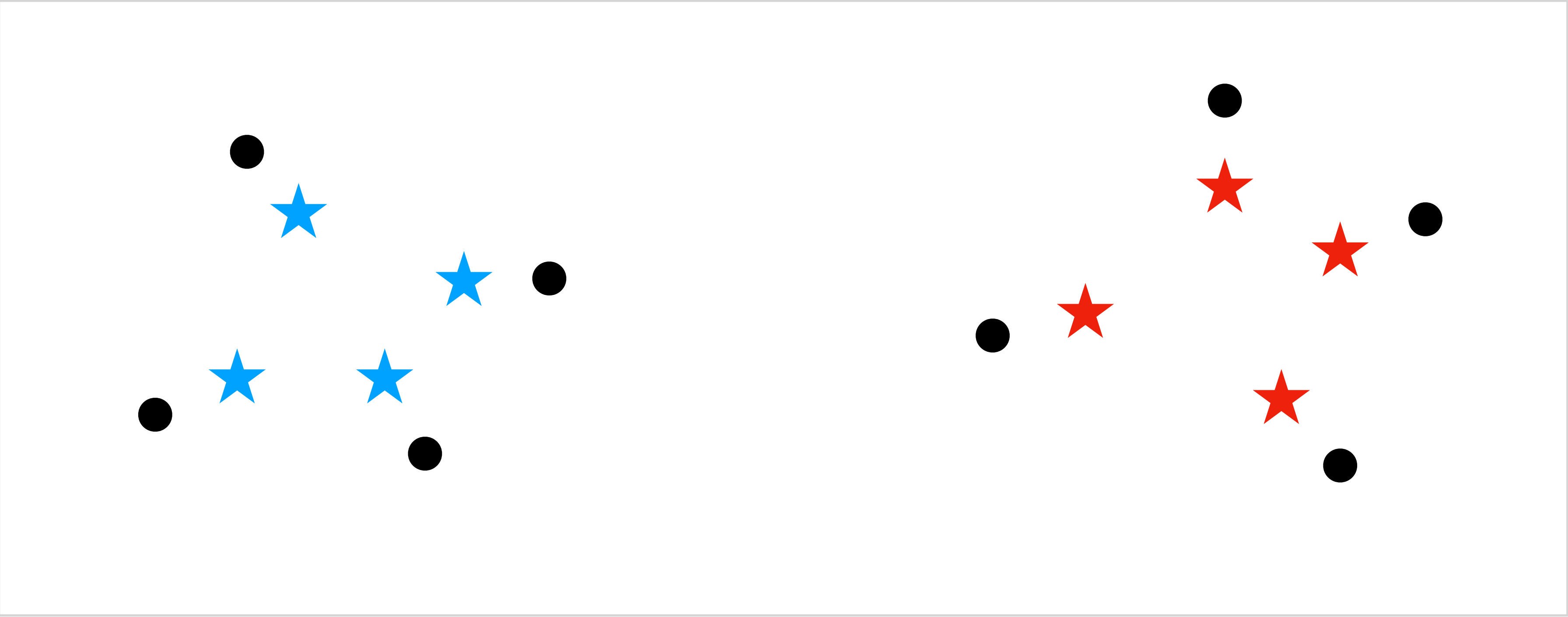
$$\underset{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{y}_i - \boldsymbol{\mu}_i\|_2^2 \quad \text{subject to} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2 \leq K'.$$



Hierarchical Clustering

A Convex Relaxation

$$\underset{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_n}{\arg\min} \quad \sum_{i=1}^{n} \|\mathbf{y}_i - \boldsymbol{\mu}_i\|_2^2 \quad \text{subject to} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2 \leq K'.$$

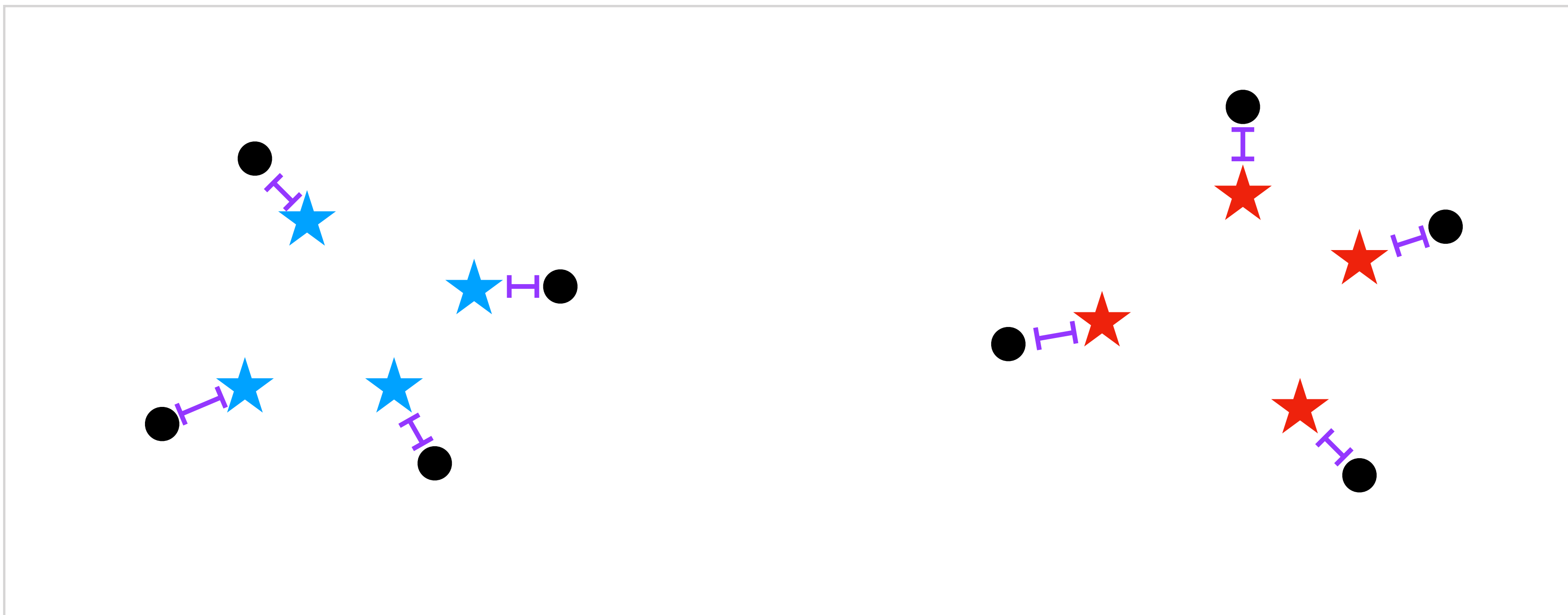


Hierarchical Clustering

A Convex Relaxation

Point-to-Center Distance

$$\underset{\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{y}_i - \boldsymbol{\mu}_i\|_2^2 \quad \text{subject to} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2 \leq K'.$$



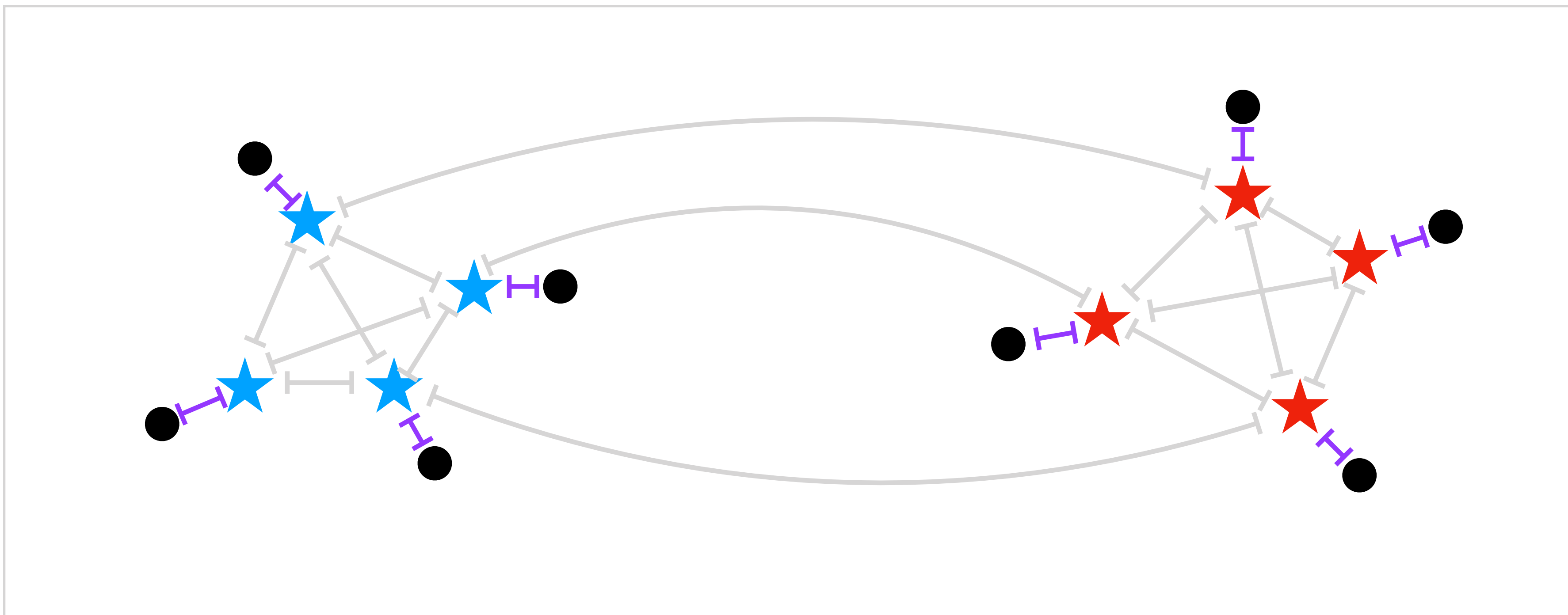
Hierarchical Clustering

A Convex Relaxation

Point-to-Center Distance

Center-to-Center Distance

$$\underset{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{y}_i - \boldsymbol{\mu}_i\|_2^2 \quad \text{subject to} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2 \leq K'.$$



Hierarchical Clustering

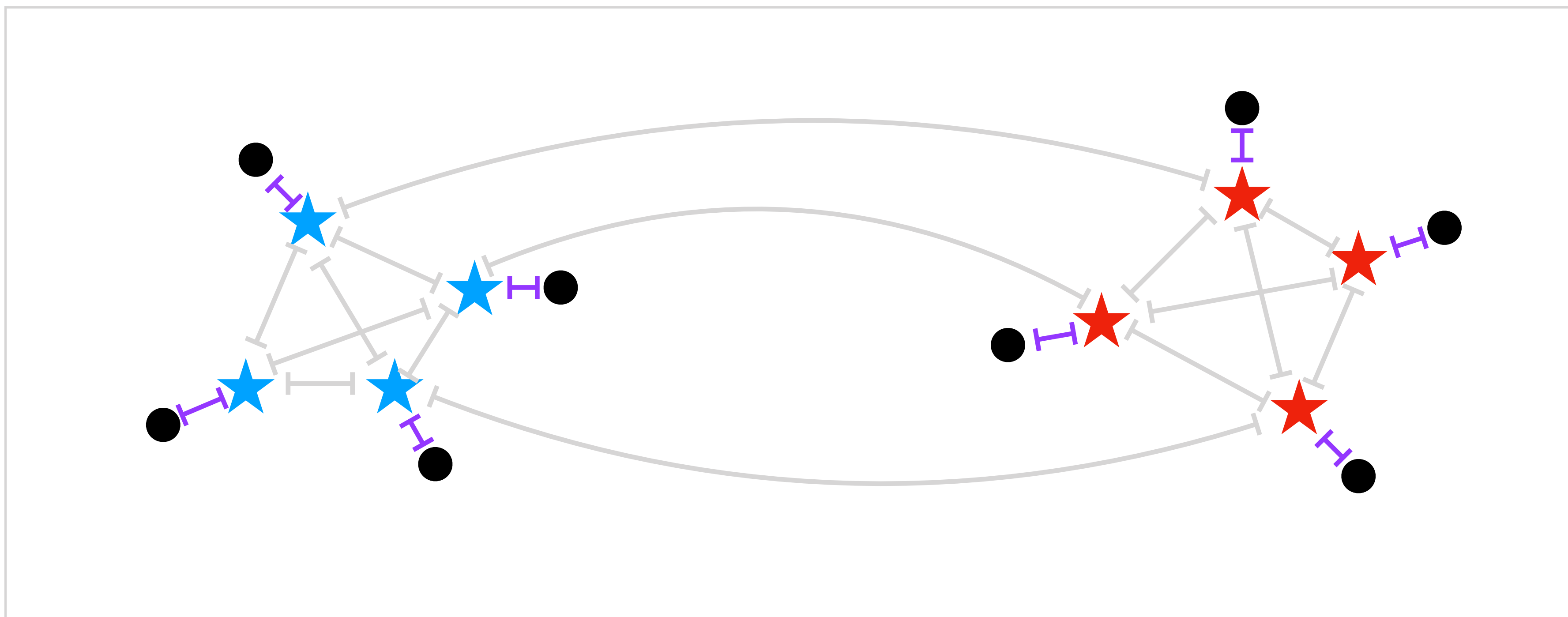
A Convex Relaxation

Point-to-Center Distance

Center-to-Center Distance

Tolerance

$$\underset{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{y}_i - \boldsymbol{\mu}_i\|_2^2 \quad \text{subject to} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2 \leq K'.$$



Hierarchical Clustering

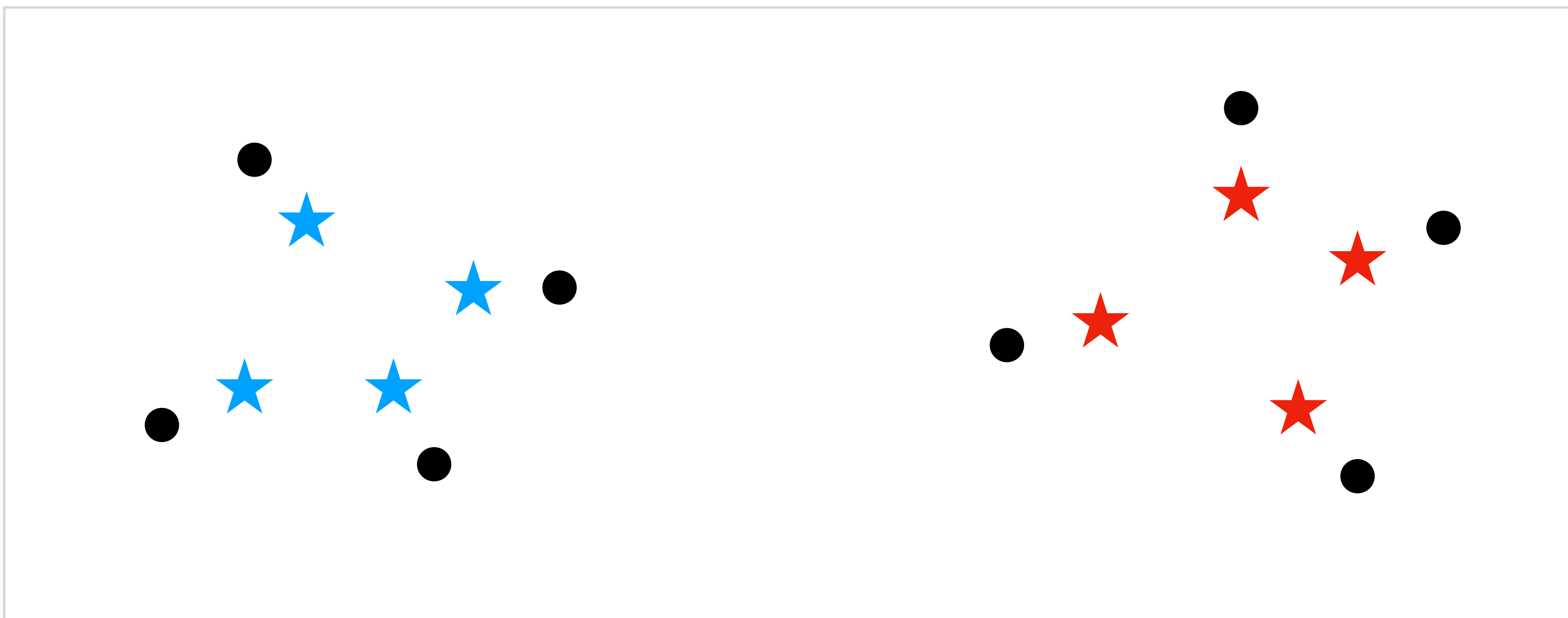
A Convex Relaxation

Point-to-Center Distance

Center-to-Center Distance

Tolerance

$$\arg\min_{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_n} \sum_{i=1}^{n} \|\mathbf{y}_i - \boldsymbol{\mu}_i\|_2^2 \quad \text{subject to} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2 \leq K'.$$



Hierarchical Clustering

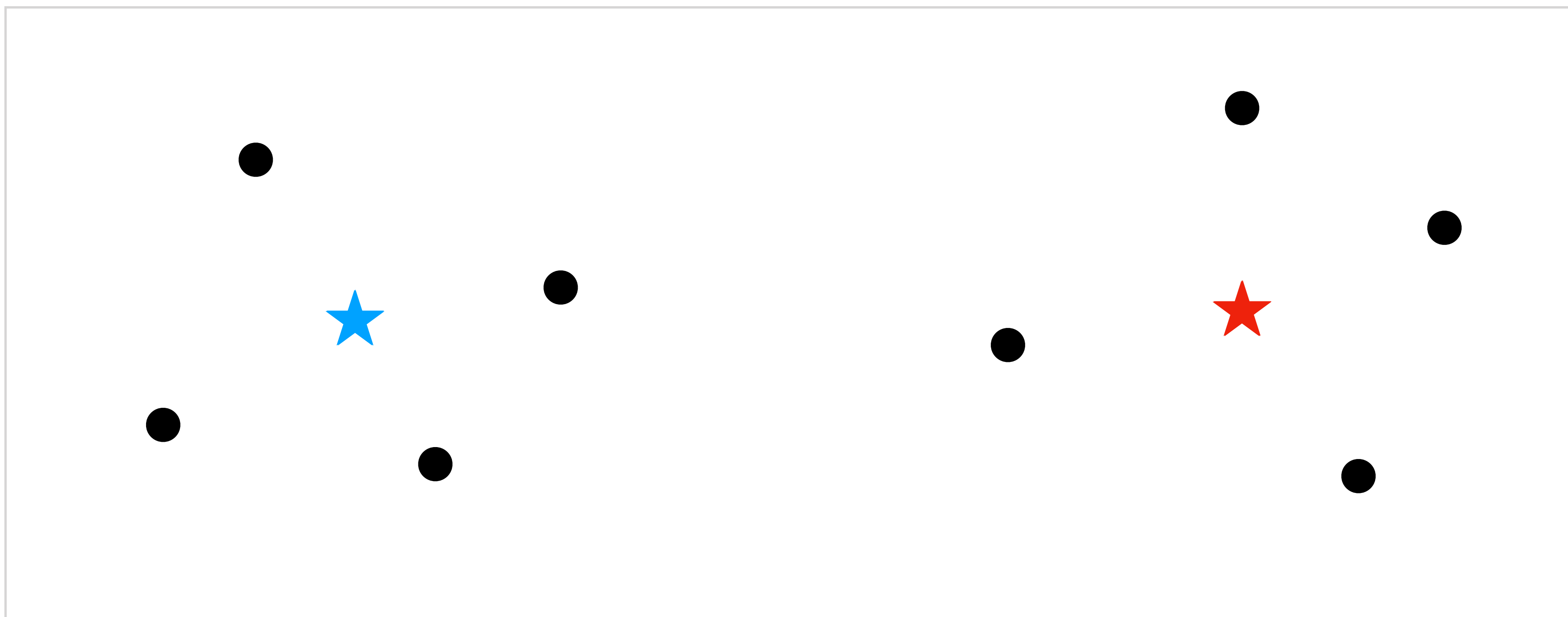
A Convex Relaxation

Point-to-Center Distance

Center-to-Center Distance

Tolerance

$$\arg\min_{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_n} \sum_{i=1}^{n} \|\mathbf{y}_i - \boldsymbol{\mu}_i\|_2^2 \quad \text{subject to} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2 \leq K'.$$



Hierarchical Clustering

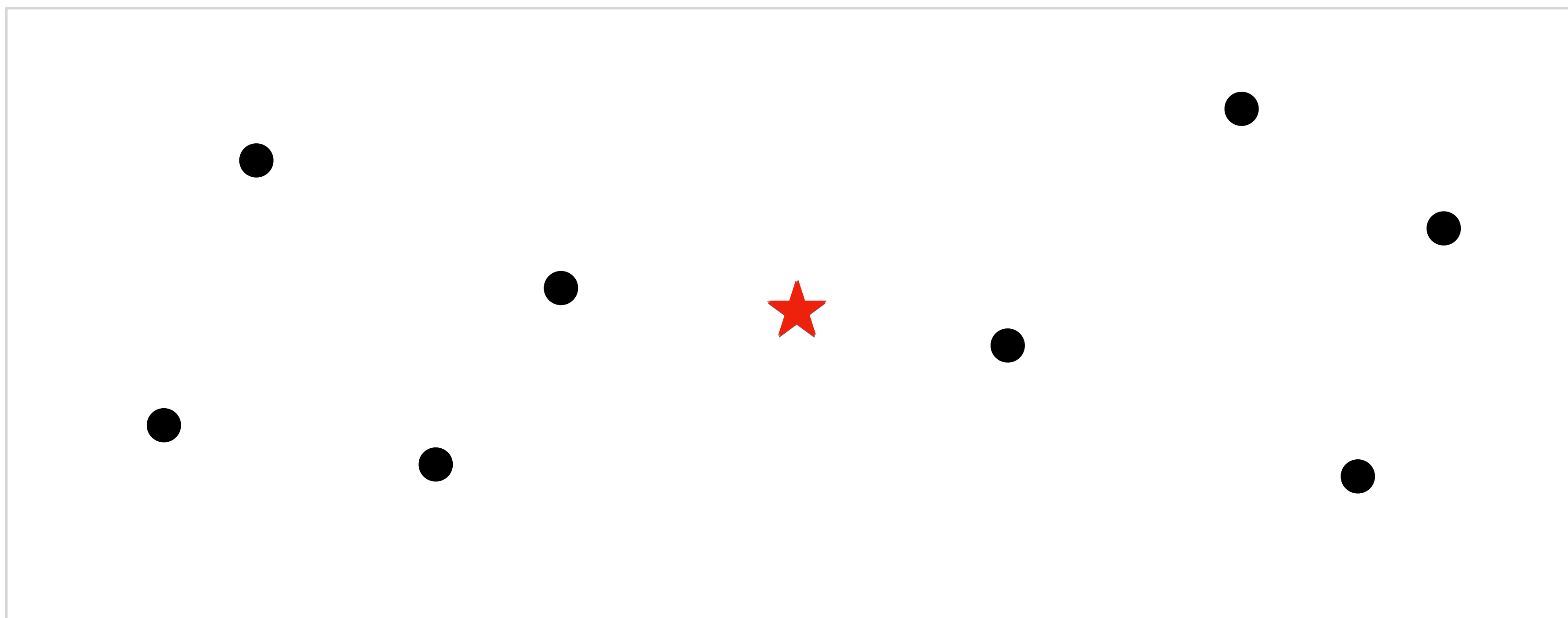
A Convex Relaxation

Point-to-Center Distance

Center-to-Center Distance

Tolerance

$$\underset{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{y}_i - \boldsymbol{\mu}_i\|_2^2 \quad \text{subject to} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2 \leq K'.$$



Hierarchical Clustering

A Convex Relaxation

Standard Clustering

$$\underset{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{y}_i - \boldsymbol{\mu}_i\|_2^2 \quad \text{subject to} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2 \leq K'.$$

Subspace Clustering

$$\underset{\mathbf{U}_1,\ldots,\mathbf{U}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{P}_i\mathbf{x}_i\|_2^2 + \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$

$$\mathbf{P}_i := \mathbf{U}_i(\mathbf{U}_i^\mathsf{T}\mathbf{U}_i)^{-1}\mathbf{U}_i^\mathsf{T}$$

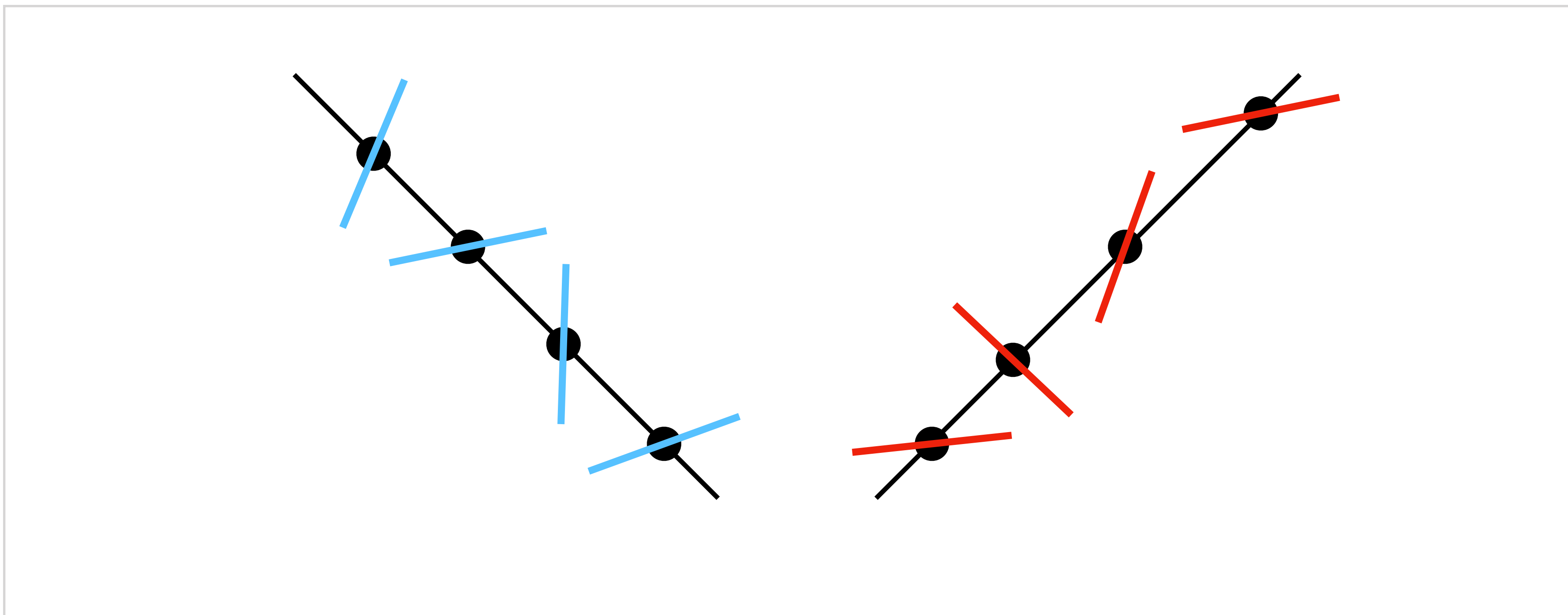
Fusion Subspace Clustering

Same idea, but for Subspaces

Point-to-Subspace Distance

Subspace-to-Subspace Distance

$$\arg\min_{\mathbf{U}_1,\ldots,\mathbf{U}_n} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{P}_i \mathbf{x}_i\|_2^2 + \frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$



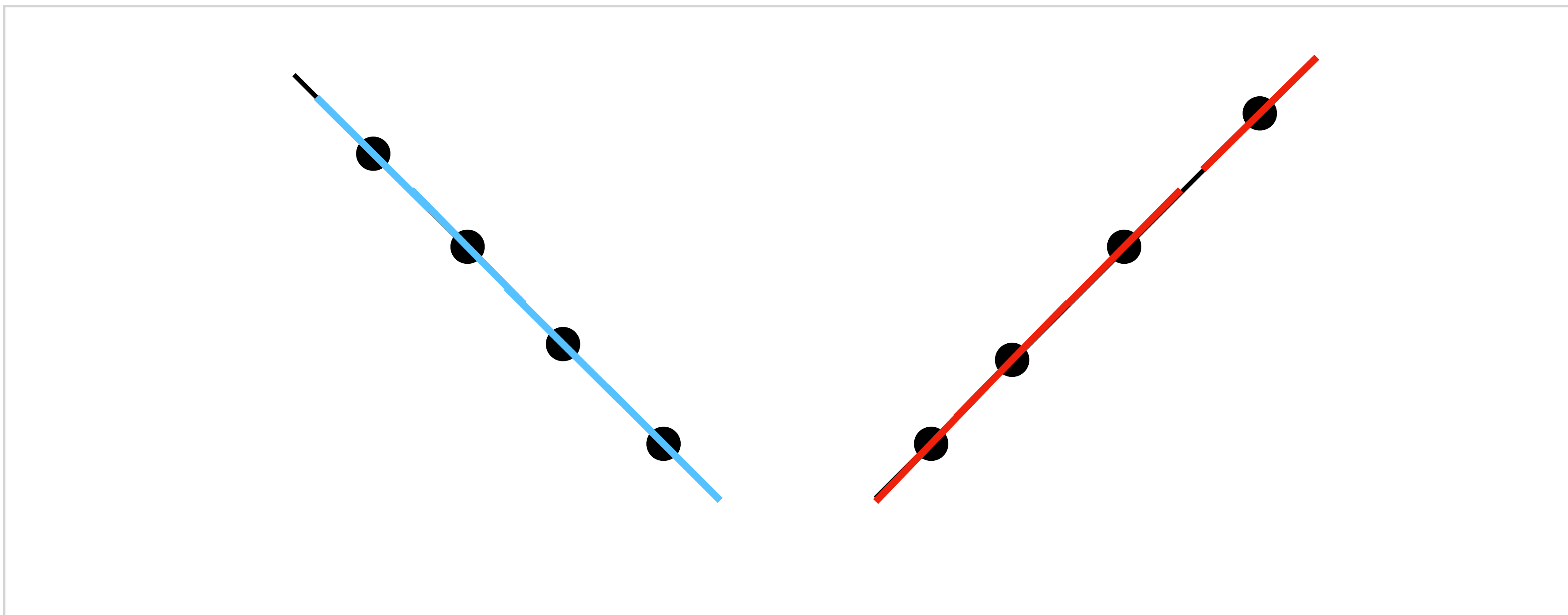
Fusion Subspace Clustering

As λ grows…

Point-to-Subspace Distance

Subspace-to-Subspace Distance

$$\arg\min_{\mathbf{U}_1,\ldots,\mathbf{U}_n} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{P}_i\mathbf{x}_i\|_2^2 + \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$



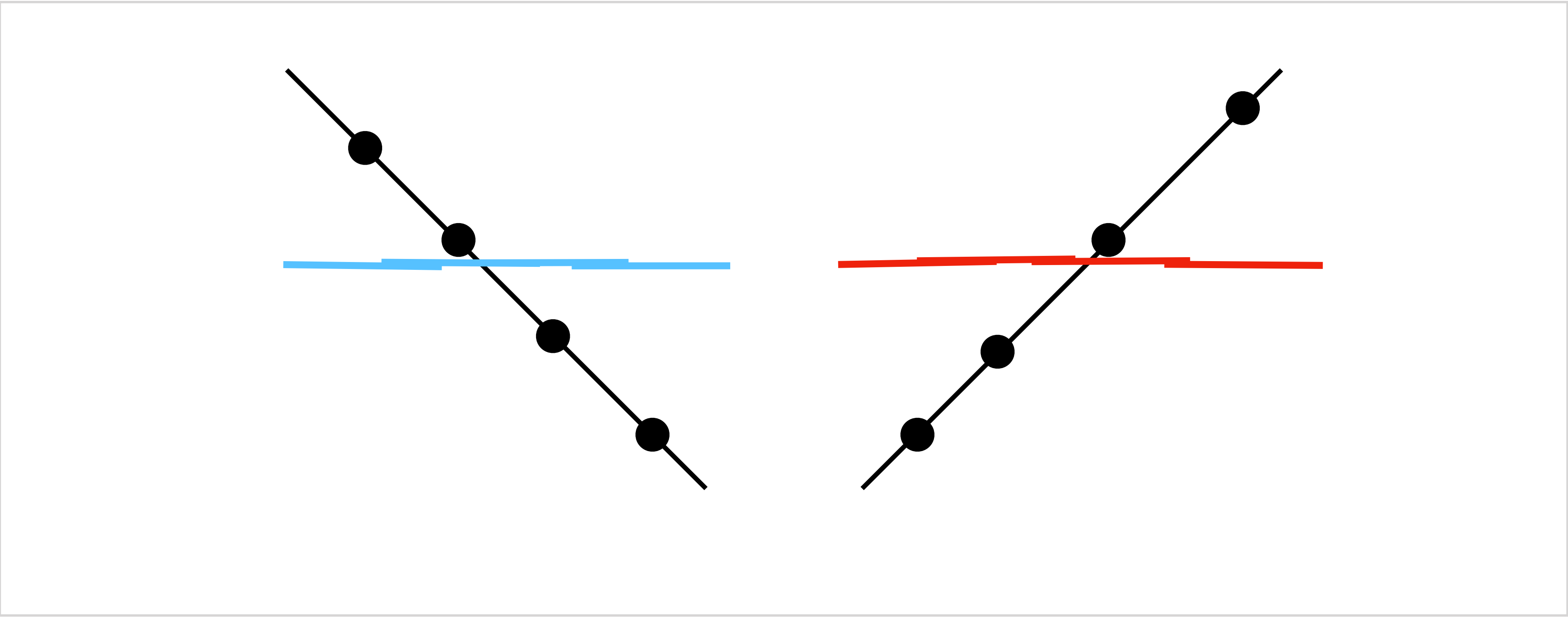
Fusion Subspace Clustering

As λ grows…

Point-to-Subspace Distance

Subspace-to-Subspace Distance

$$\underset{\mathbf{U}_1,\ldots,\mathbf{U}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{P}_i \mathbf{x}_i\|_2^2 + \frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$



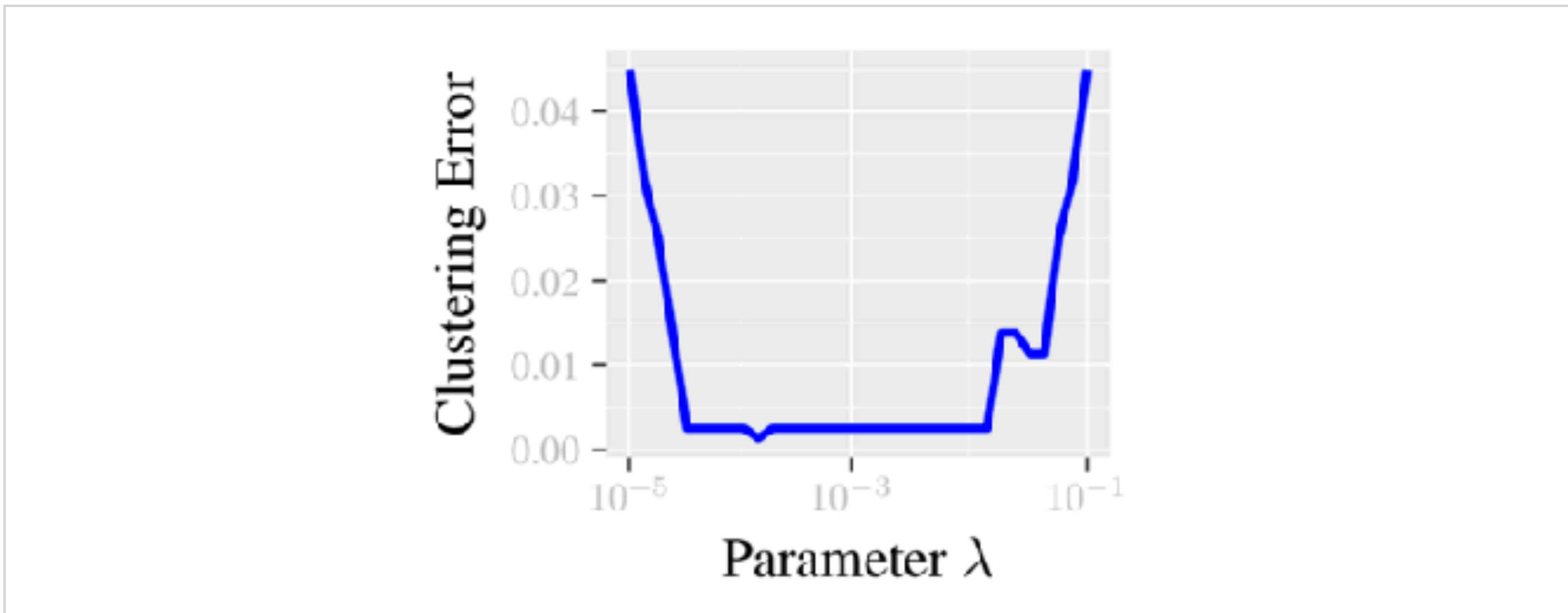
Fusion Subspace Clustering

As λ grows…

Point-to-Subspace Distance

Subspace-to-Subspace Distance

$$\arg\min_{\mathbf{U}_1,\ldots,\mathbf{U}_n} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{P}_i\mathbf{x}_i\|_2^2 + \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$



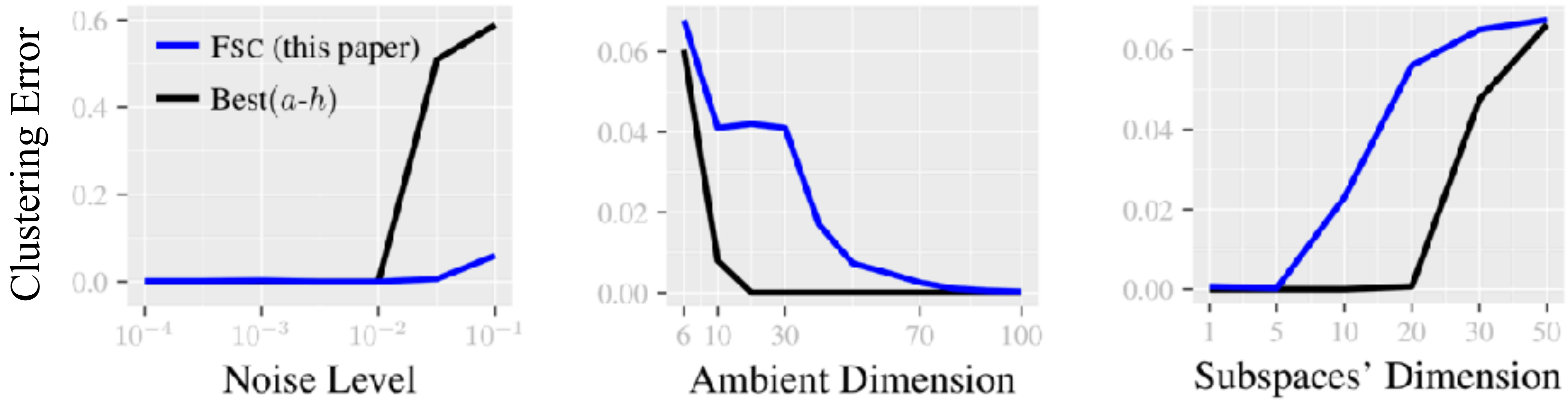
Fusion Subspace Clustering

As λ grows…

$$\arg\min_{\mathbf{U}_1,\ldots,\mathbf{U}_n} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{P}_i\mathbf{x}_i\|_2^2 + \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$

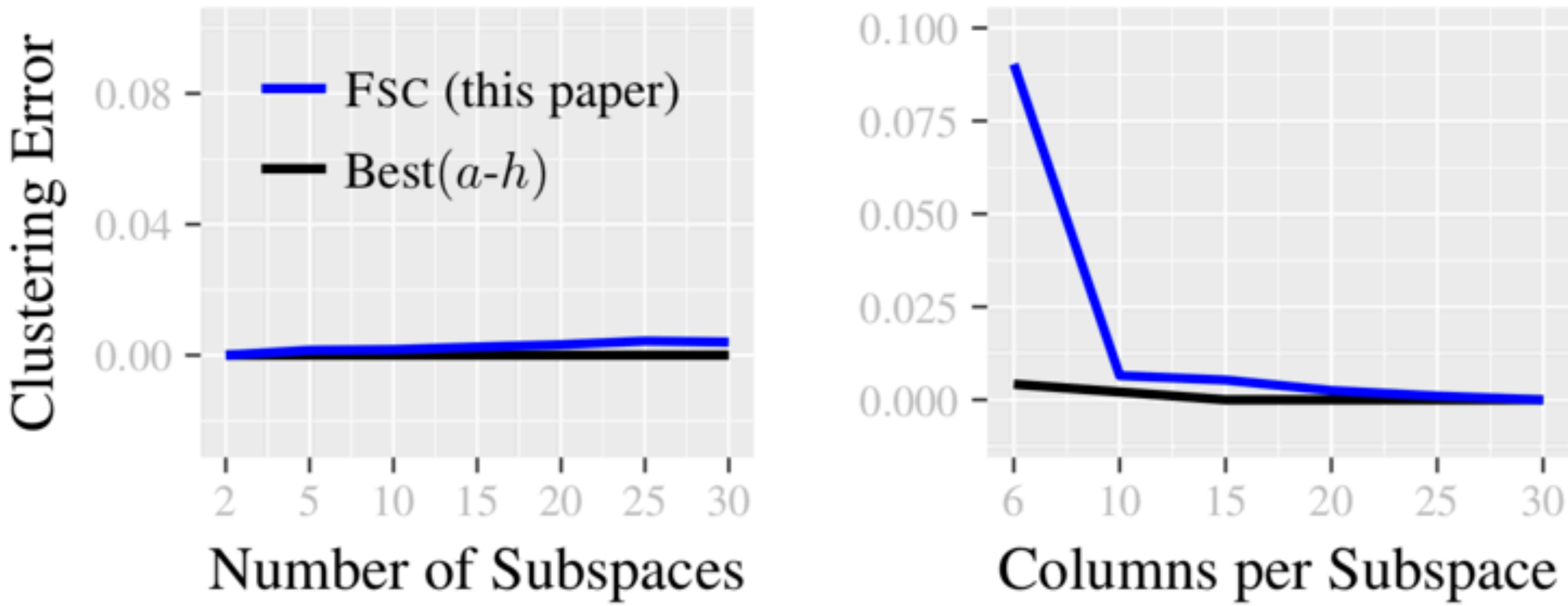
Fusion Subspace Clustering

Initialization



Fusion Subspace Clustering

A Few Experiments



Fusion Subspace Clustering

A Few Experiments


Why do I care?
SSC seems
better…

Why do I care?
SSC seems
better…

The Answer
is Missing

Why do I care?
SSC seems
better…

The Answer
is Missing …
Missing Data

Motion Segmentation

Incomplete Data often lies in a Union of Subspaces

Motion Segmentation

Incomplete Data often lies in a Union of Subspaces

Motion Segmentation

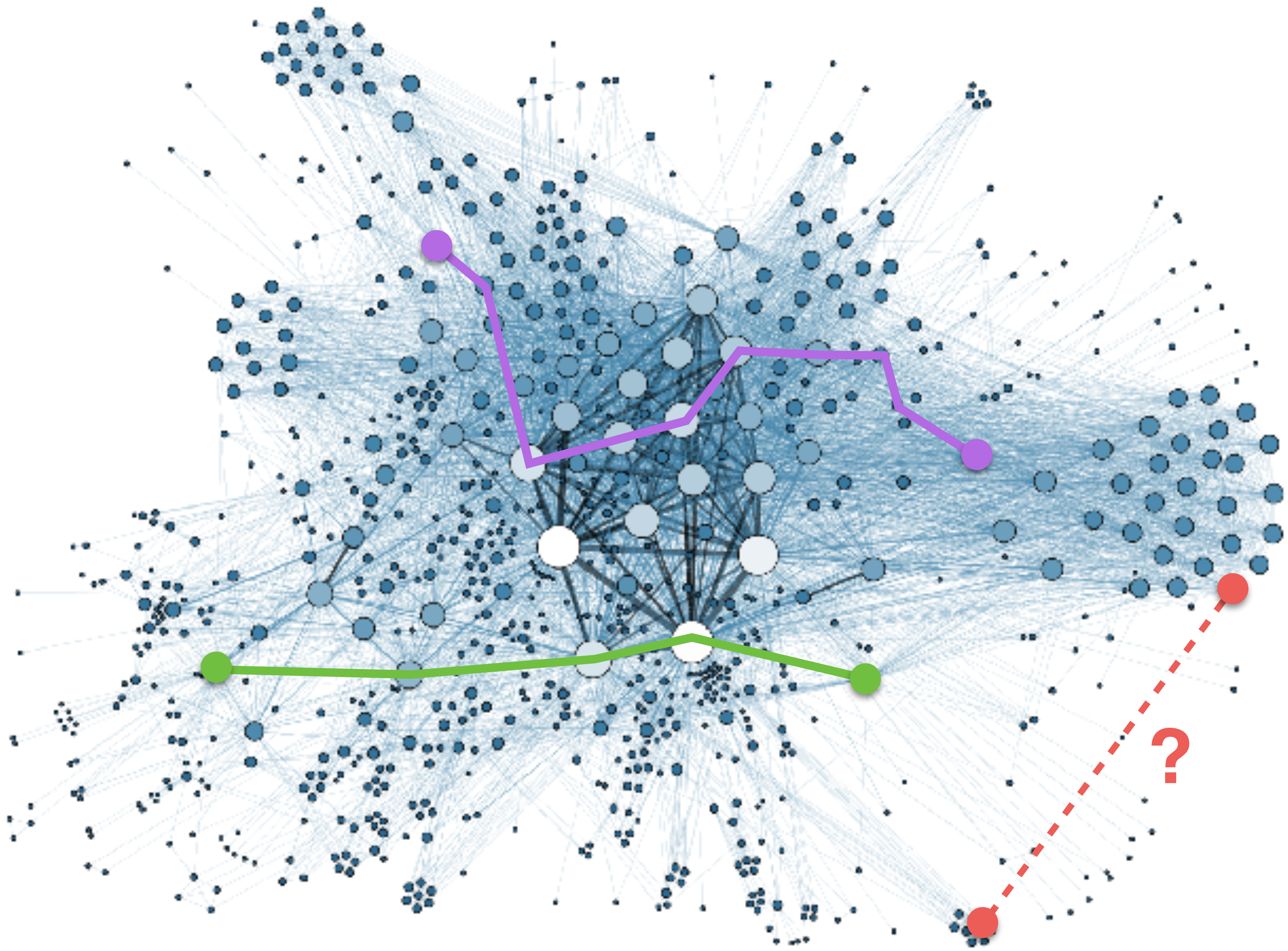
Incomplete Data often lies in a Union of Subspaces

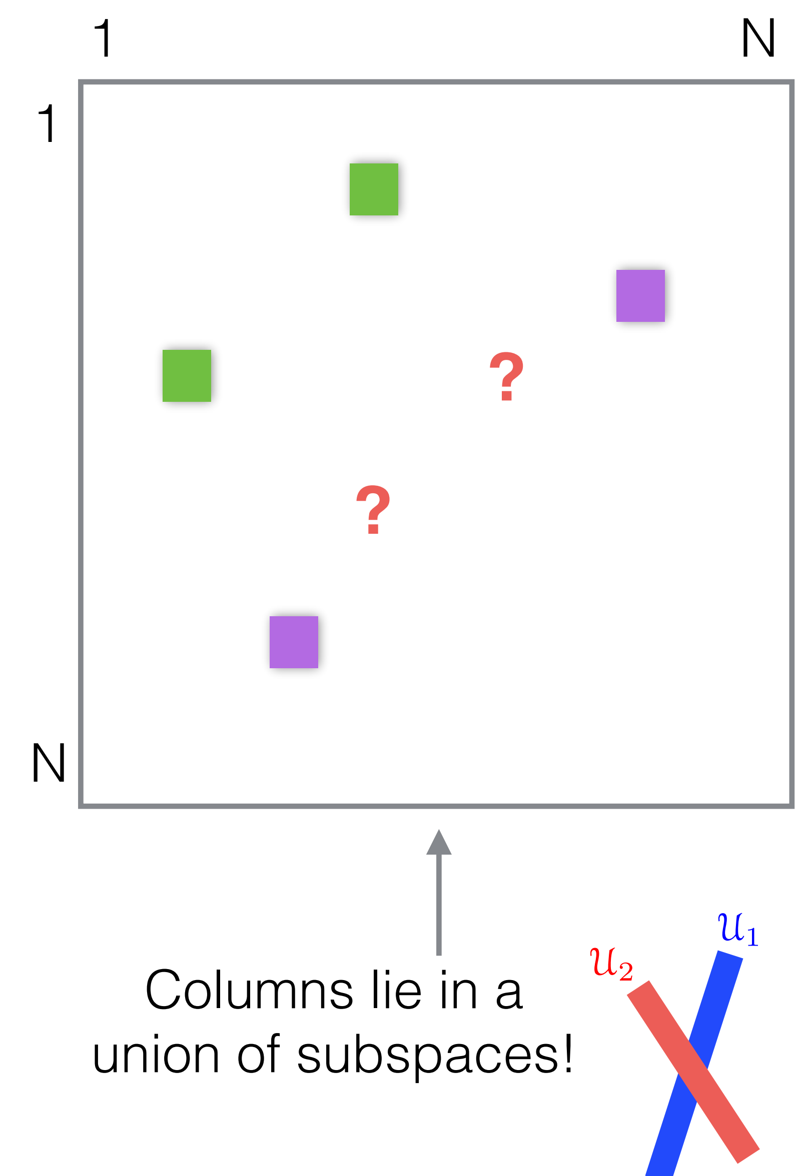
Motion Segmentation

Incomplete Data often lies in a Union of Subspaces

Motion Segmentation

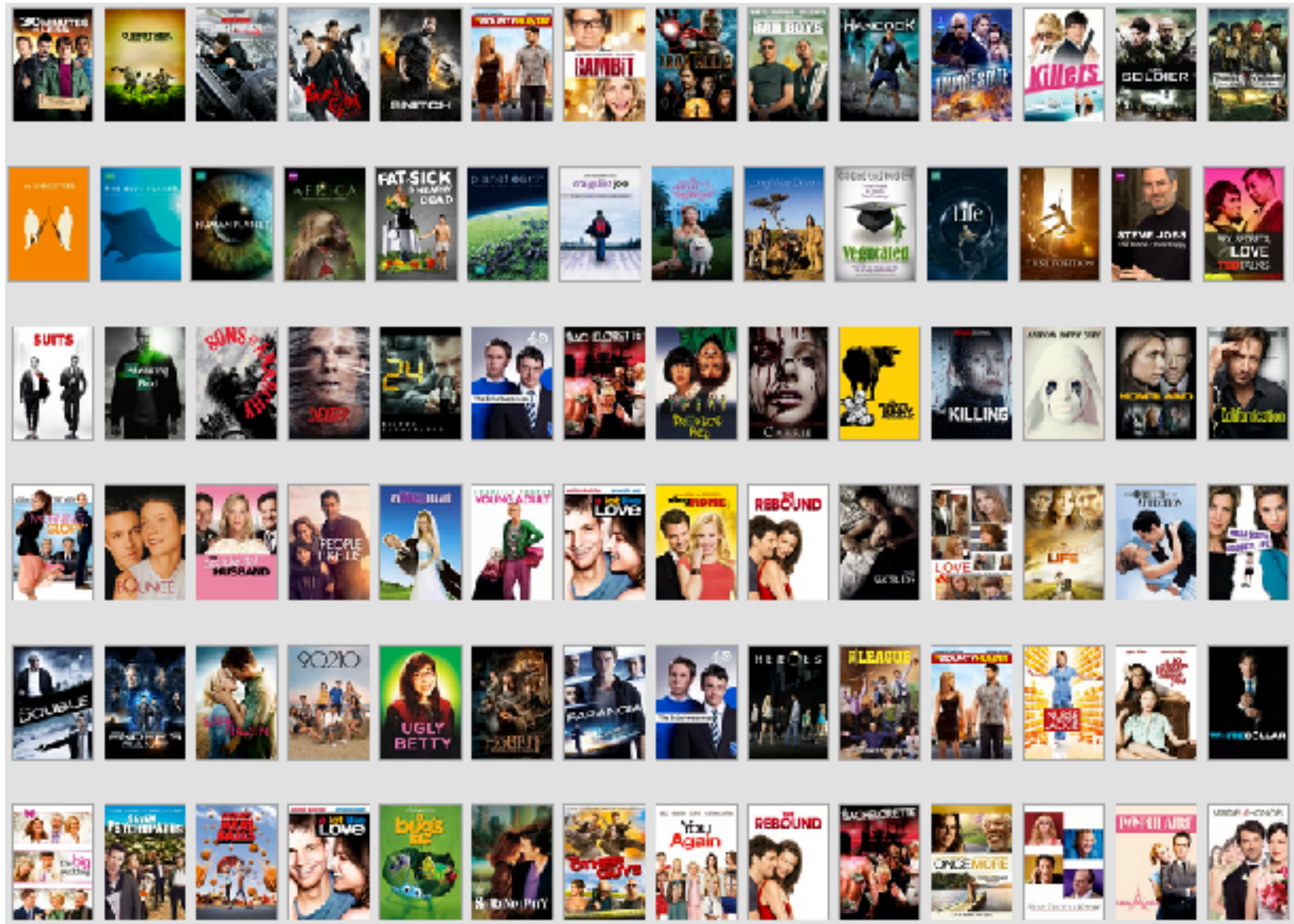
Incomplete Data often lies in a Union of Subspaces





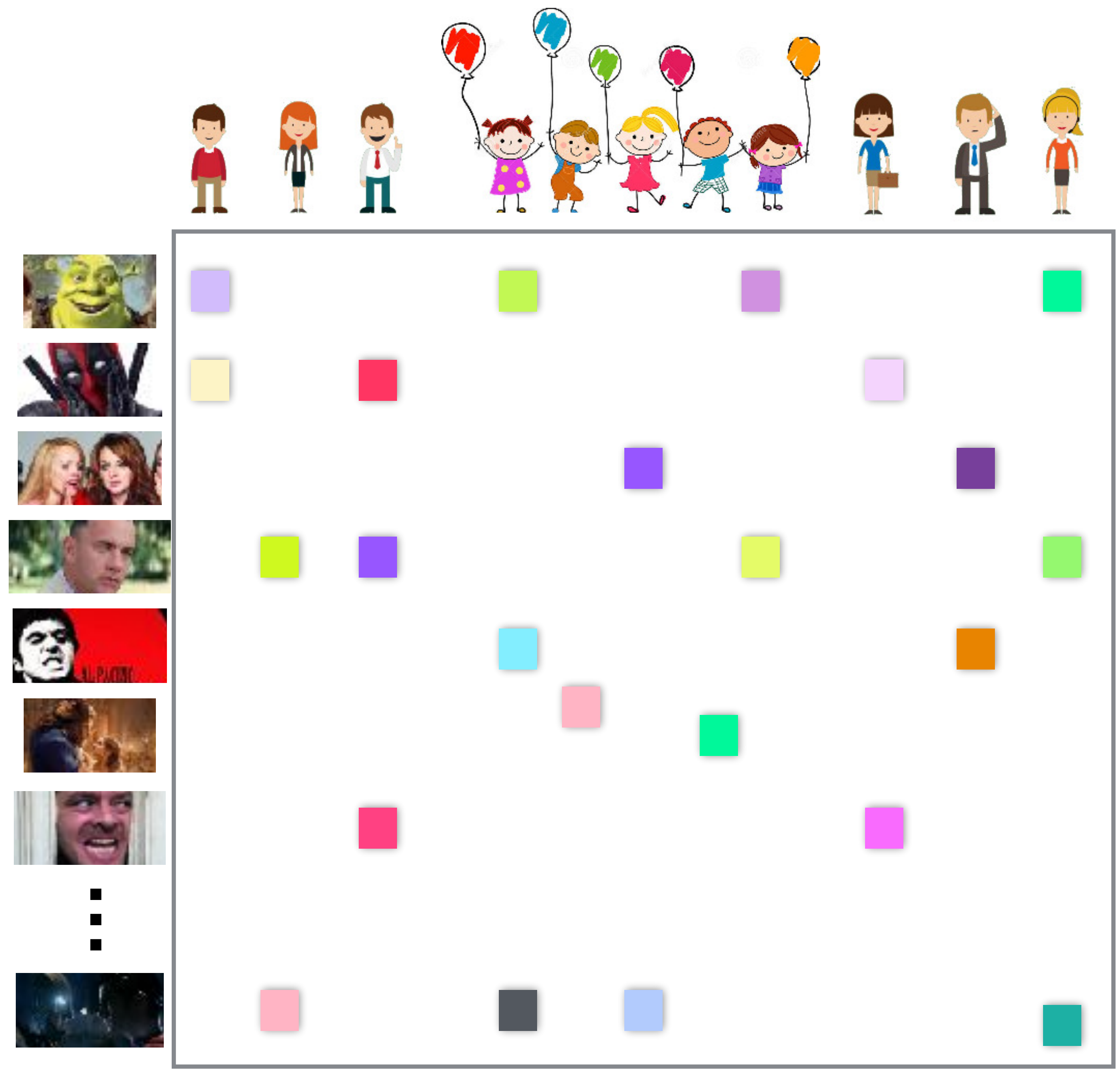
Network Estimation

Incomplete Data often lies in a Union of Subspaces



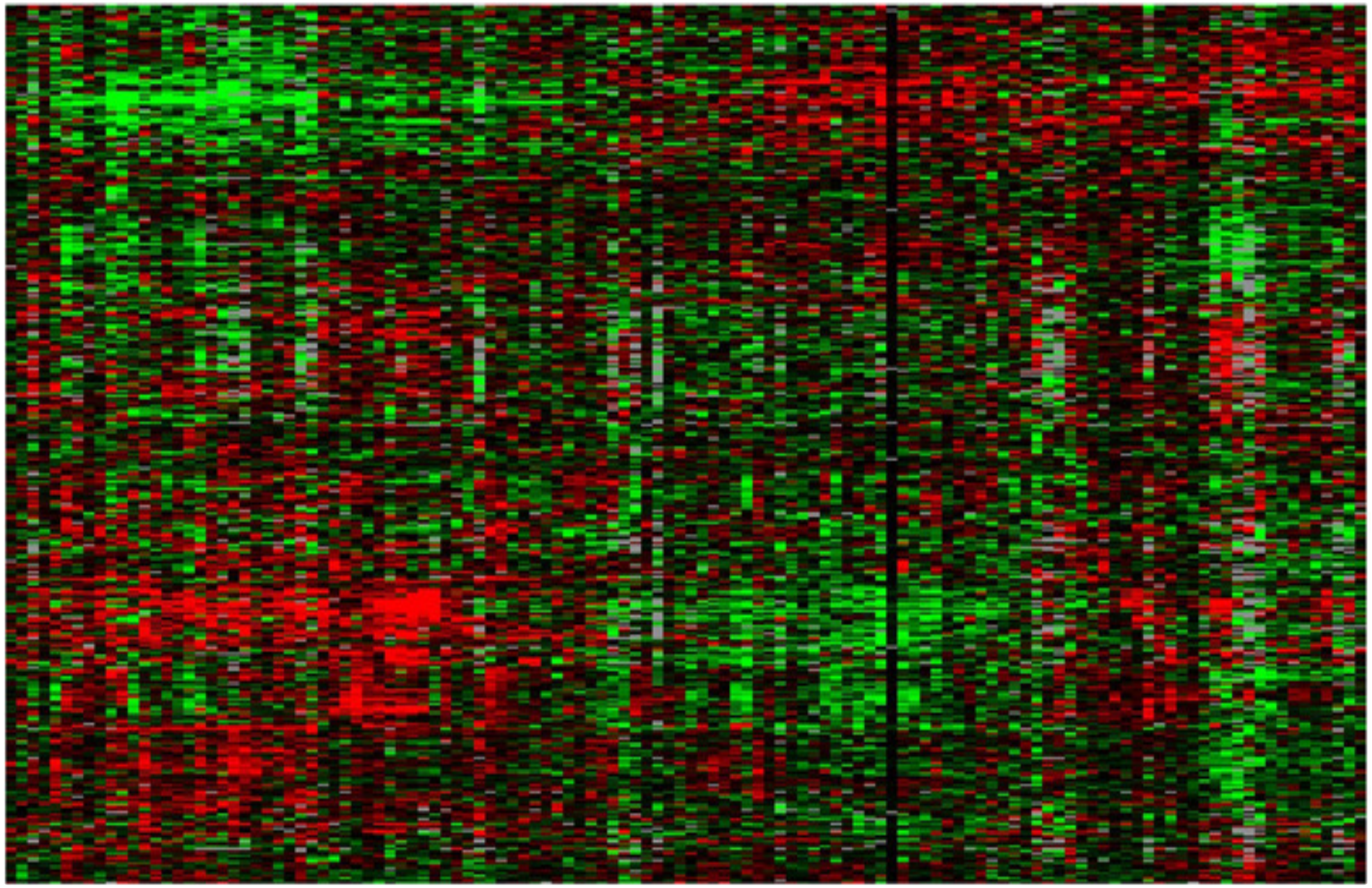
Recommender Systems

Incomplete Data often lies in a Union of Subspaces



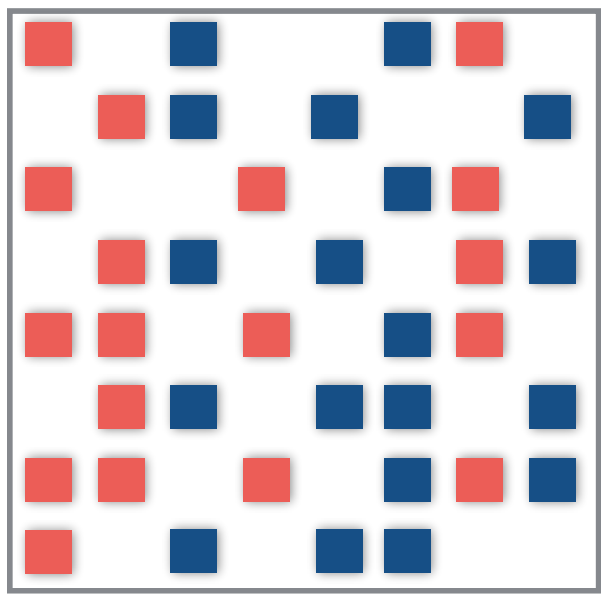
Recommender Systems

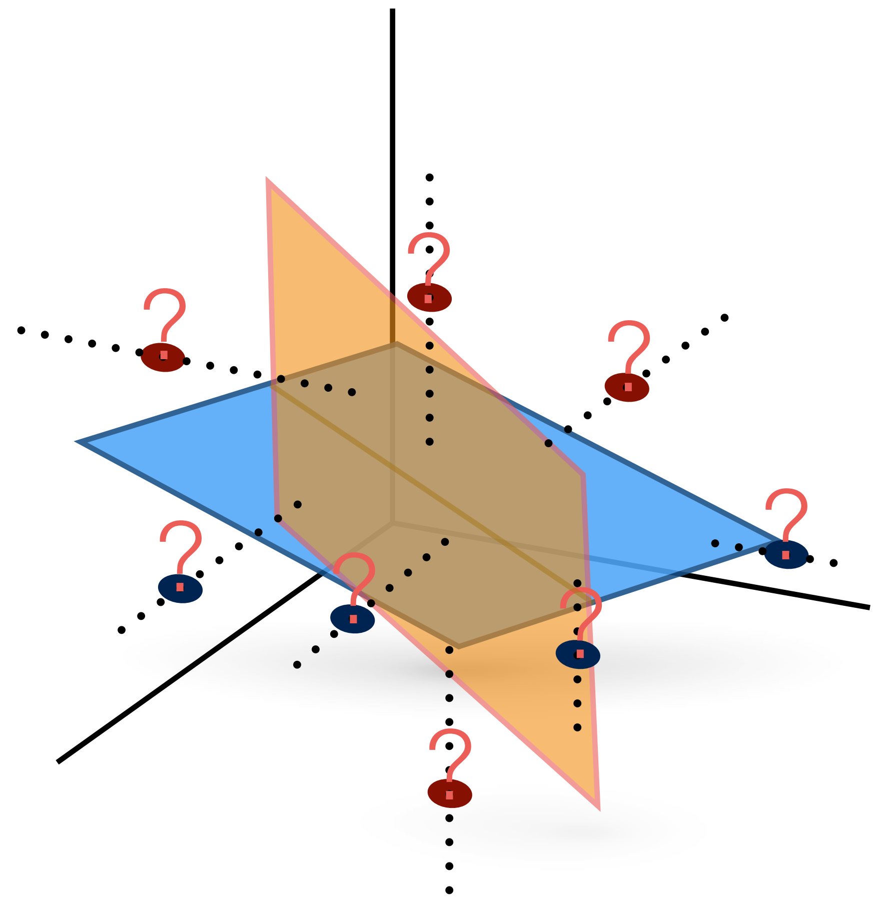
Incomplete Data often lies in a Union of Subspaces



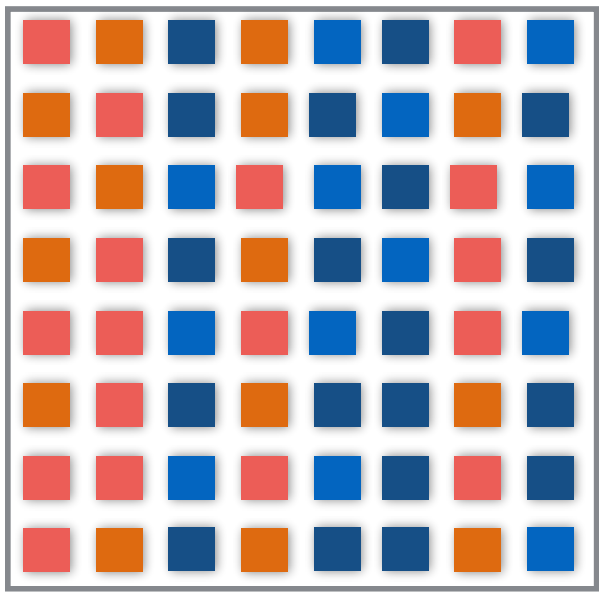
Genomics & Drug Discovery

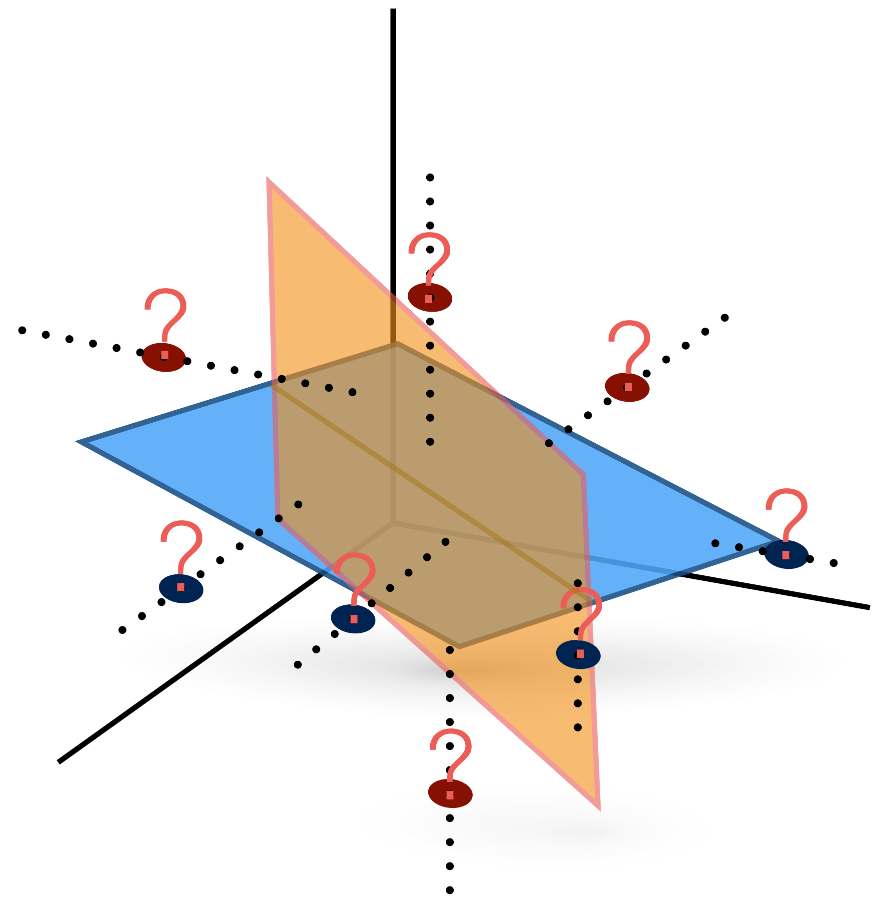
Incomplete Data often lies in a Union of Subspaces





Subspace Clustering with Missing Data.



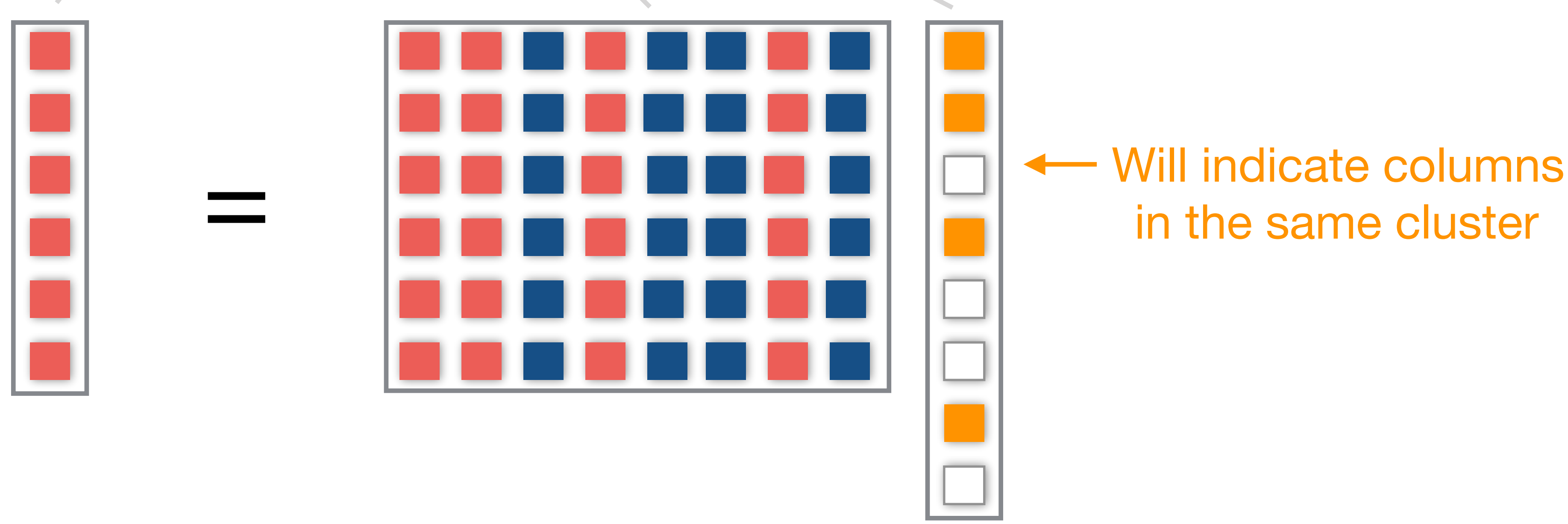


Subspace Clustering with Missing Data.

a.k.a. High-Rank Matrix Completion


Why not just use SSC?

$$\min_{\boldsymbol{\theta}_i} \ \|\mathbf{x}_i - \mathbf{X}_{-i}\boldsymbol{\theta}_i\|_2 \ + \ \frac{\lambda}{2}\|\boldsymbol{\theta}_i\|_1$$

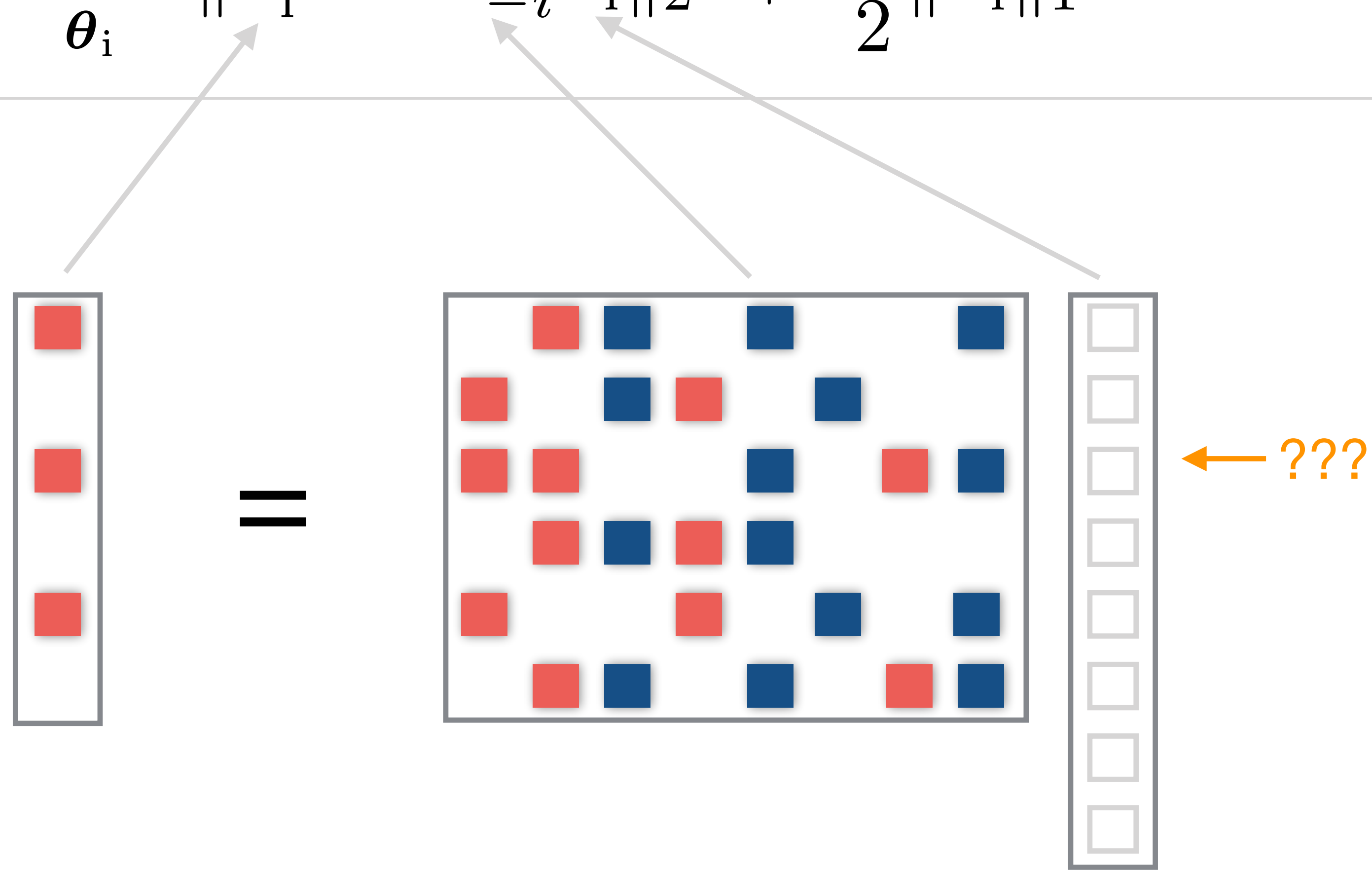


Sparse Subspace Clustering

Main idea: Write each column as a **sparse** linear combination

Works Fantastically with Full Data

$$\min_{\boldsymbol{\theta}_\mathrm{i}} \ \|\mathbf{x}_\mathrm{i}^{\omega} - \mathbf{X}_{-i}^{\Omega}\boldsymbol{\theta}_\mathrm{i}\|_2 \ + \ \frac{\lambda}{2}\|\boldsymbol{\theta}_\mathrm{i}\|_1$$



Sparse Subspace Clustering

Main idea: Write each column as a **sparse** linear combination

Doesn't Work if Data is Missing

ALGORITHMS	Advantage	Disadvantage
EWZF-SSC	Works with little Missing Data	Fails with little Missing Data
LRMC+SSC	Works with Low-Rank Data	Fails unless Low-Rank Data
Group Sparse SC	State-of-the-art	Local Minima & Initialization
EM	State-of-the-art	Local Minima & Initialization
Liftings	Works with Small Data	Fails with Big Data

Subspace Clustering w/ Missing Data

Alternatives to SSC

Fusion Subspace Clustering

$$\underset{\mathbf{U}_1,\ldots,\mathbf{U}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{P}_i \mathbf{x}_i\|_2^2 + \frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$

Fusion Subspace Clustering with Missing Data

Incomplete Data

Full Subspace

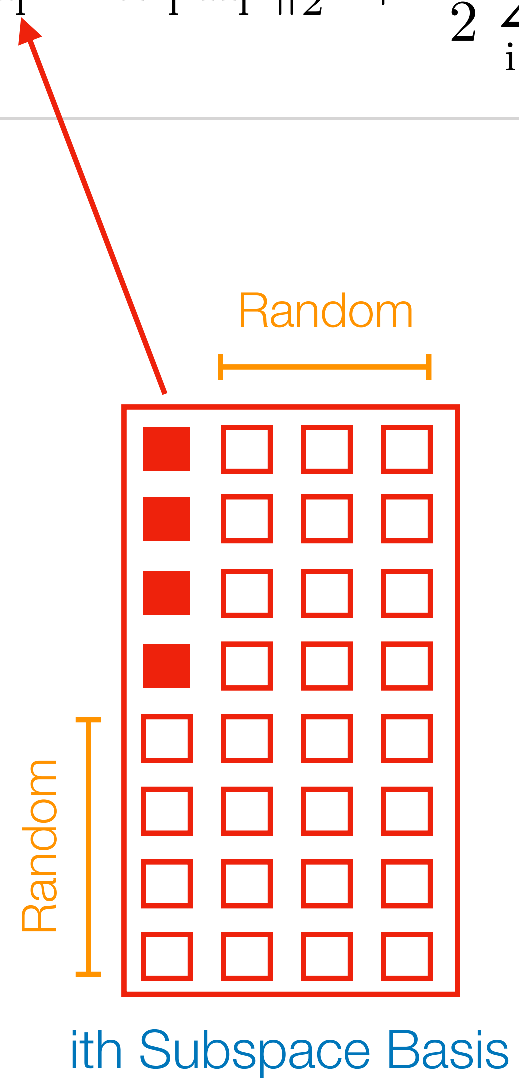
$$\underset{\mathbf{U}_1,\ldots,\mathbf{U}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{x}_i^{\omega} - \mathbf{P}_i^{\omega} \mathbf{x}_i^{\omega}\|_2^2 + \frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$

$$\mathbf{P}_i^{\omega} := \mathbf{U}_i^{\omega} (\mathbf{U}_i^{\omega \mathsf{T}} \mathbf{U}_i^{\omega})^{-1} \mathbf{U}_i^{\omega \mathsf{T}}$$

Fusion Subspace Clustering

Natural Extension to Missing Data

$$\underset{\mathbf{U}_1,\ldots,\mathbf{U}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{x}_i^{\omega} - \mathbf{P}_i^{\omega}\mathbf{x}_i^{\omega}\|_2^2 + \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$

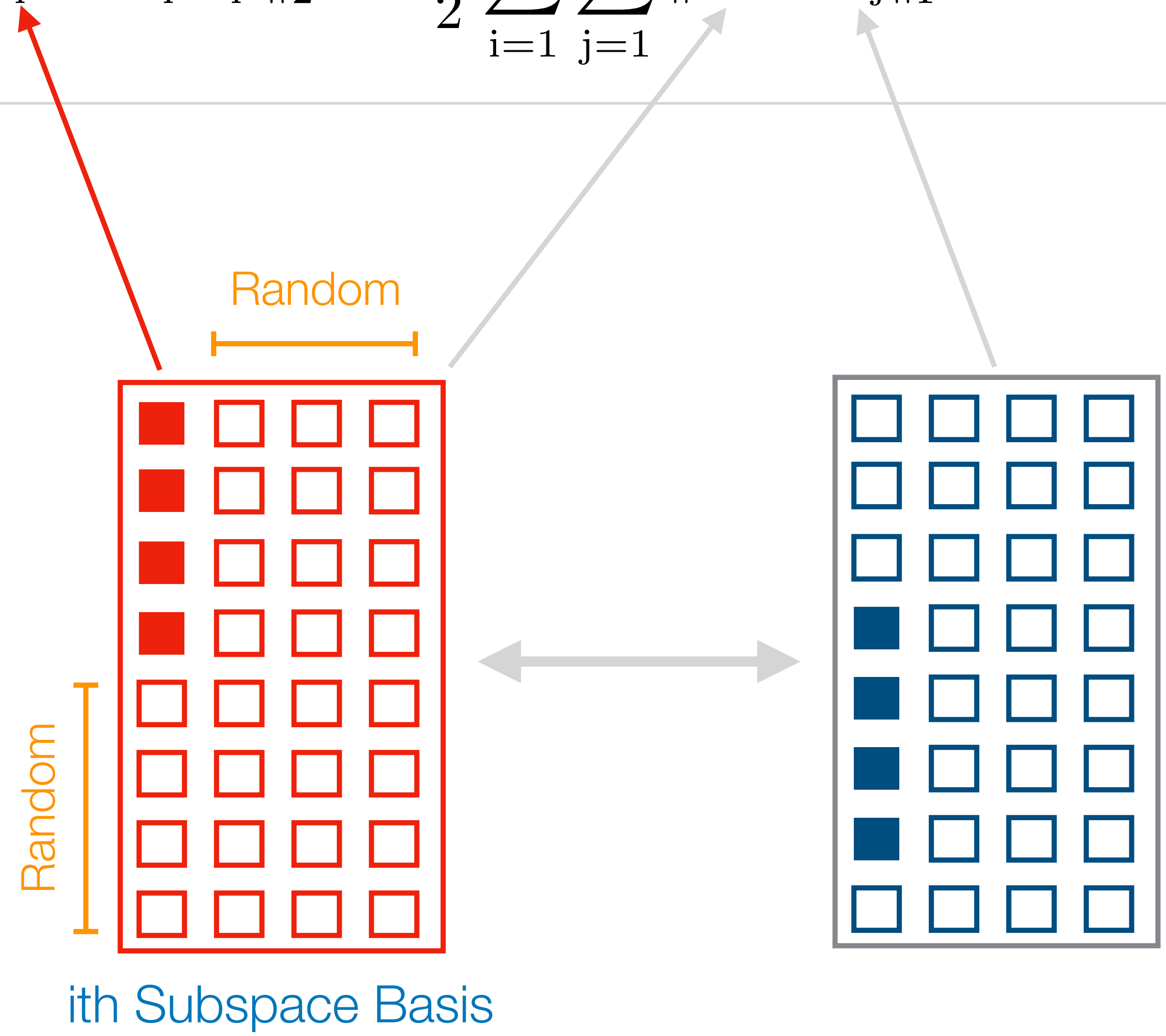


Fusion Subspace Clustering

Natural Extension to Missing Data

Initialization

$$\underset{\mathbf{U}_1,\ldots,\mathbf{U}_n}{\arg\min} \sum_{i=1}^{n} \|\mathbf{x}_i^{\omega} - \mathbf{P}_i^{\omega}\mathbf{x}_i^{\omega}\|_2^2 + \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$



Fusion Subspace Clustering

Natural Extension to Missing Data

Initialization

$$\underset{\mathbf{U}_1,\ldots,\mathbf{U}_n}{\arg\min} \quad \sum_{i=1}^{n} \|\mathbf{x}_i^{\omega} - \mathbf{P}_i^{\omega}\mathbf{x}_i^{\omega}\|_2^2 \; + \; \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{P}_i - \mathbf{P}_j\|_F^2$$

ith column coefficient:

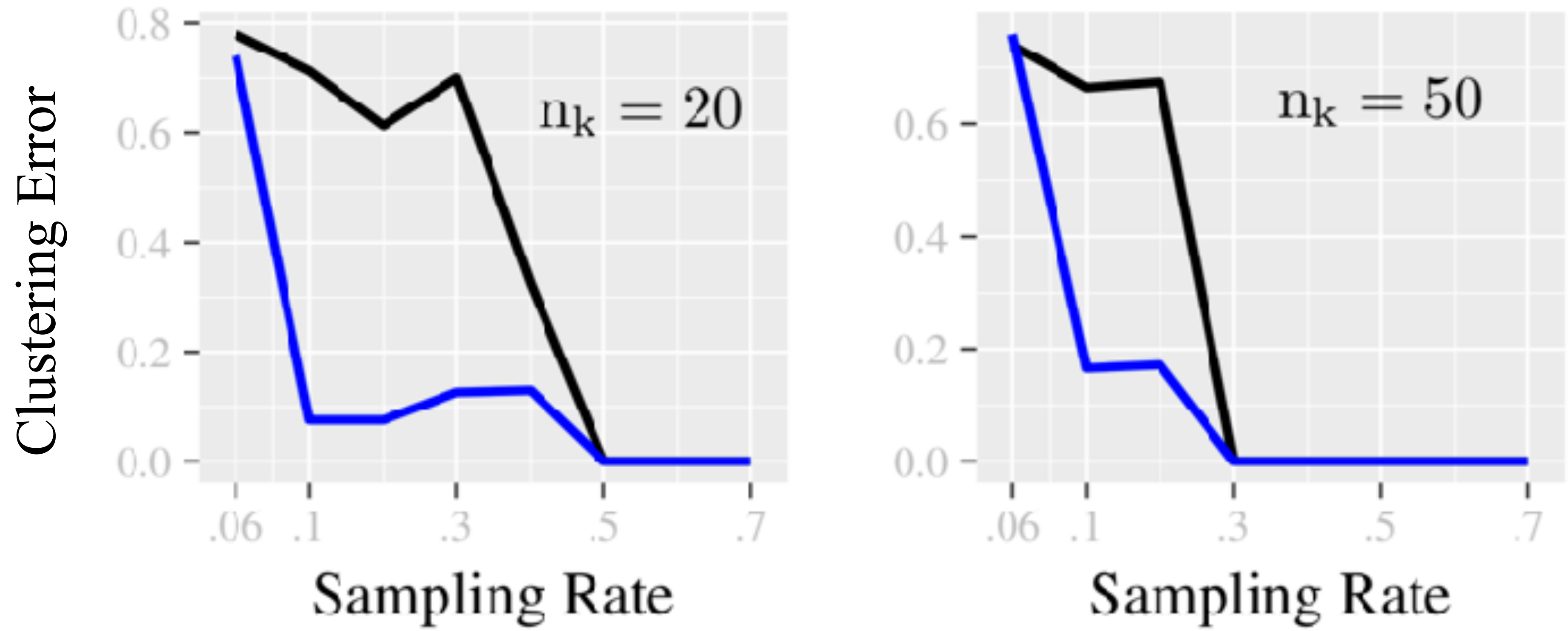
$$\hat{\boldsymbol{\theta}}_i := (\hat{\mathbf{U}}_{k_i}^{\omega\mathsf{T}}\hat{\mathbf{U}}_{k_i}^{\omega})^{-1}\hat{\mathbf{U}}_{k_i}^{\omega\mathsf{T}}\mathbf{x}_i^{\omega}$$

ith column completion:

$$\hat{\mathbf{x}}_i = \hat{\mathbf{U}}_{k_i}\hat{\boldsymbol{\theta}}_i$$

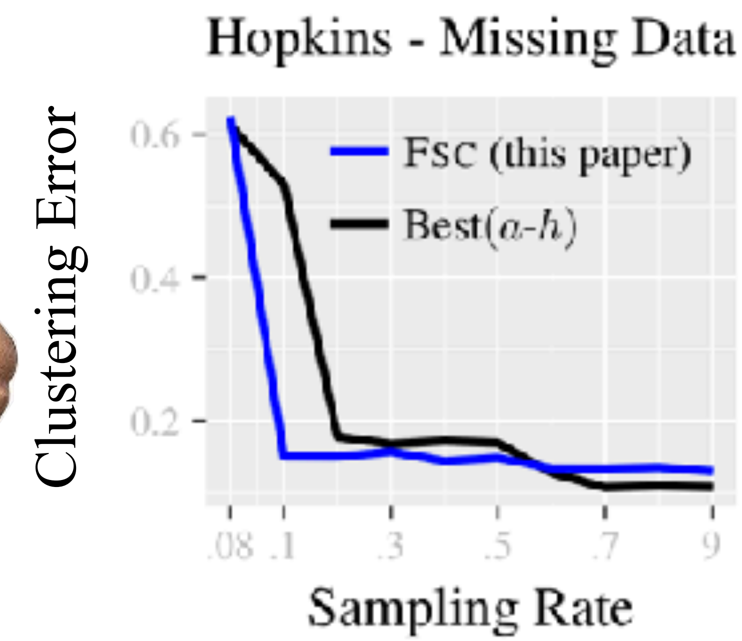
Fusion Subspace Clustering

Natural Extension to Missing Data
Completion (no need for LRMC)



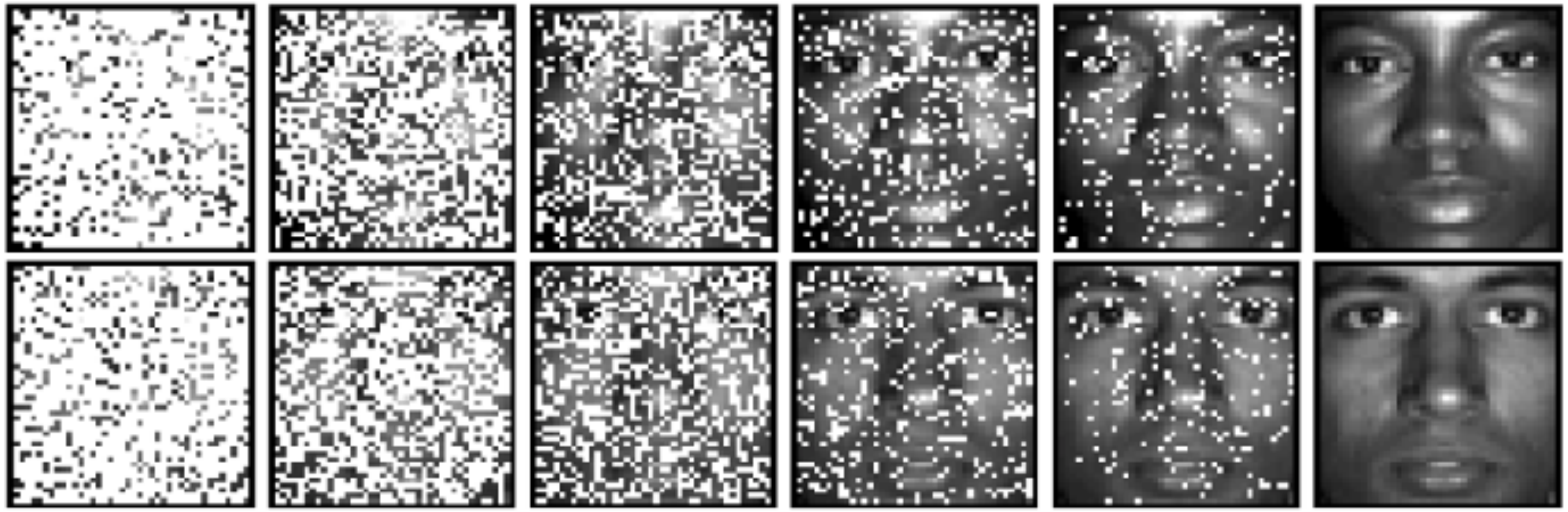
Fusion Subspace Clustering w/Missing Data

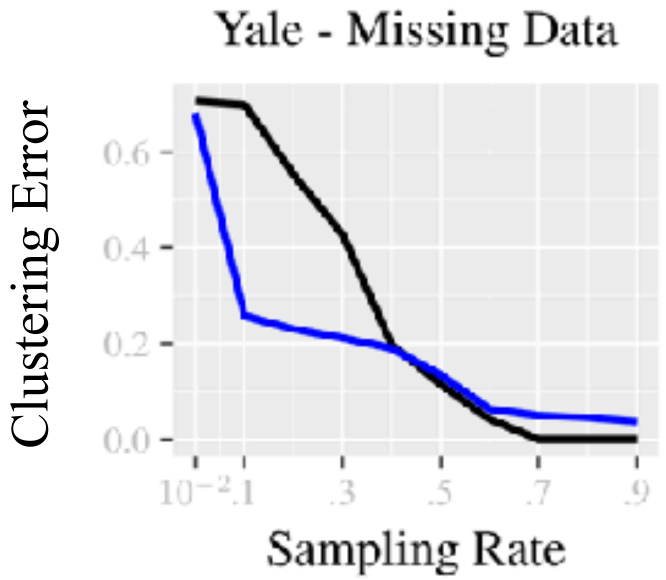
A Few Experiments



Fusion Subspace Clustering w/Missing Data

A Few Experiments





Fusion Subspace Clustering w/Missing Data

A Few Experiments

Fusion Subspace Clustering

What we know

- Competitive with full-data
- OK with noise
- Good with missing data
- Promising

Fusion Subspace Clustering

What we DON'T know (everything else)

- Sample Complexity?
- Computational Complexity?
- Convergence?
- Initialization?
- Parameters (λ)?
- Variants (greedy, adversarial, adaptive, Grassmannian)

Questions?