CS 221: Artificial Intelligence

Lecture 6:
Advanced Machine Learning

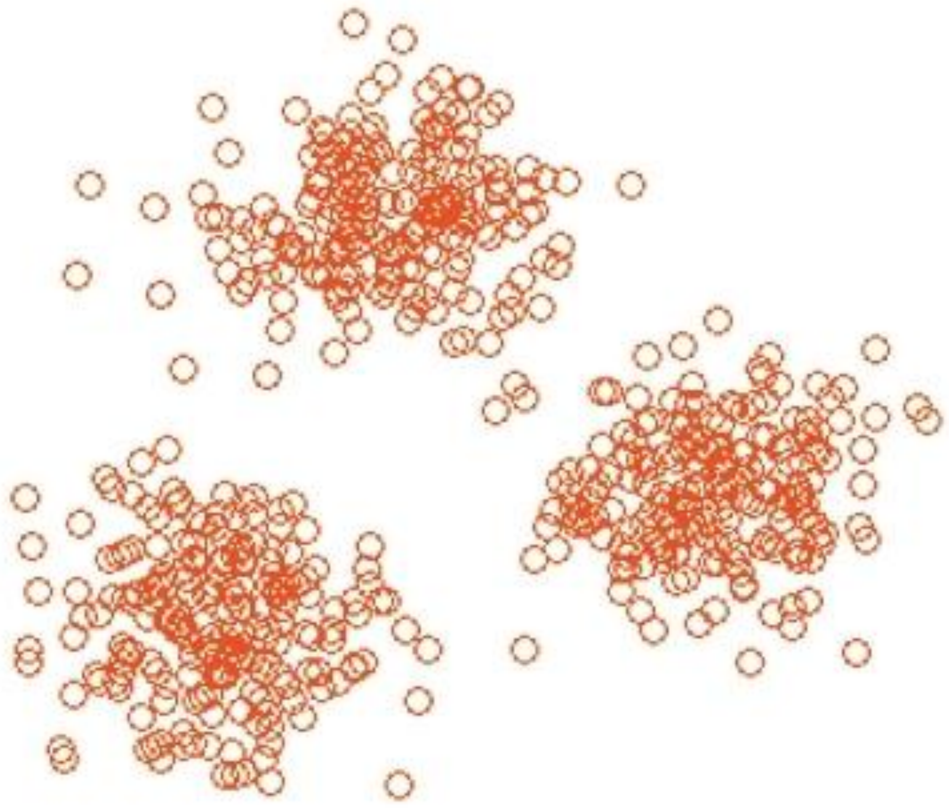
Sebastian Thrun and Peter Norvig

Slide credit: Mark Pollefeys, Dan Klein, Chris Manning

Outline

- Clustering
  - K-Means
  - EM
  - Spectral Clustering
- Dimensionality Reduction

The unsupervised learning problem

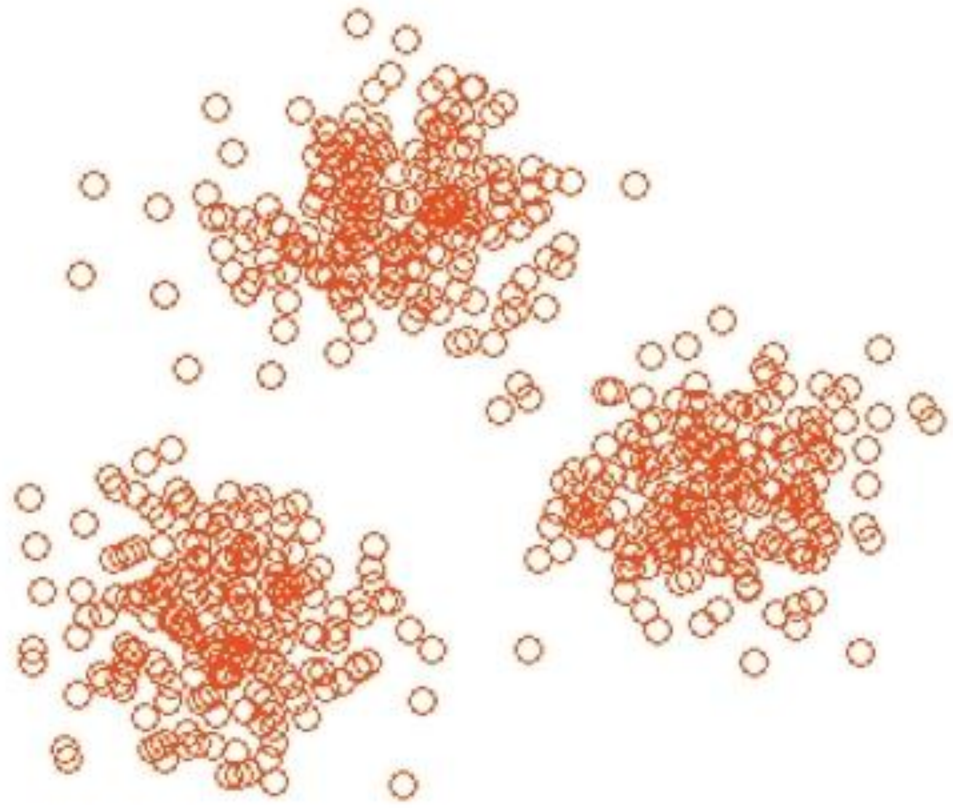


Many data points, no labels

Unsupervised Learning?

- Google Street View

K-Means



Many data points, no labels

K-Means

- Choose a fixed number of clusters

- Choose cluster centers and point-cluster allocations to minimize error
- can’t do this by exhaustive search, because there are too many possible allocations.

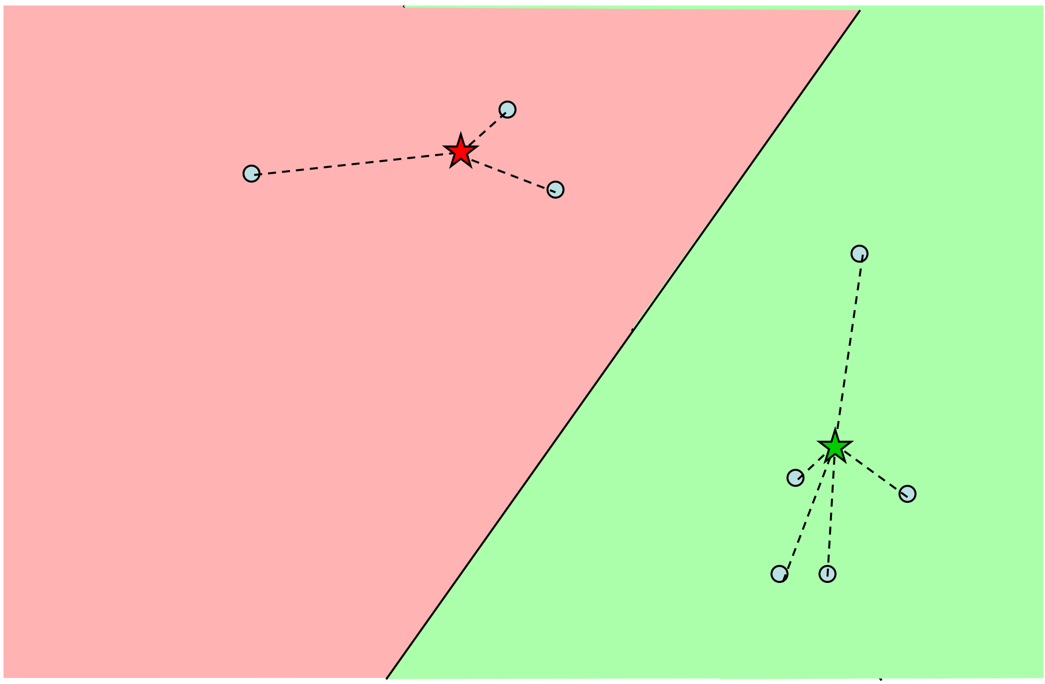
- Algorithm
  - fix cluster centers; allocate points to closest cluster
  - fix allocation; compute best cluster centers
- x could be any set of features for which we can compute a distance (careful about scaling)

$$\| \qquad \|^2$$

clusters elements of i'th cluster

* From Marc Pollefeys COMP 256 2003

K-Means



K-Means

```
Choose k data points to act as cluster centers

Until the cluster centers are unchanged

    Allocate each data point to cluster whose center is nearest

    Now ensure that every cluster has at least
    one data point; possible techniques for doing this include .
    supplying empty clusters with a point chosen at random from
    points far from their cluster center.

    Replace the cluster centers with the mean of the elements
    in their clusters.

end
```

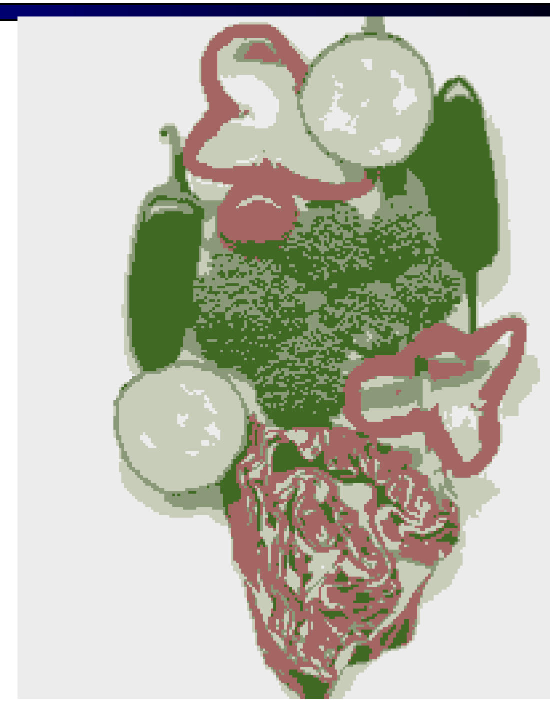
Algorithm 16.5: *Clustering by K-Means*

* From Marc Pollefeys COMP 256 2003

Results of K-Means Clustering:

Image

Clusters on intensity



Clusters on color

K-means clustering using intensity alone and color alone

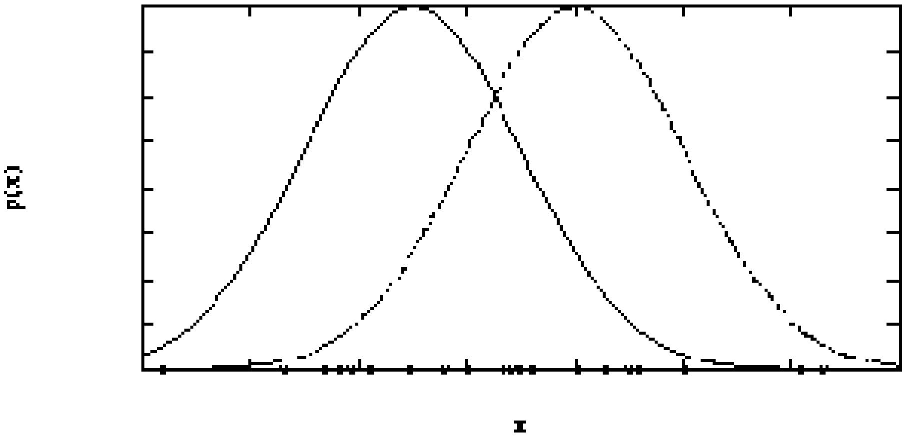
* From Marc Pollefeys COMP 256 2003

K-Means

- Is an approximation to EM
  - Model (hypothesis space): Mixture of N Gaussians
  - Latent variables: Correspondence of data and Gaussians
- We notice:
  - Given the mixture model, it's easy to calculate the correspondence
  - Given the correspondence it's easy to estimate the mixture models

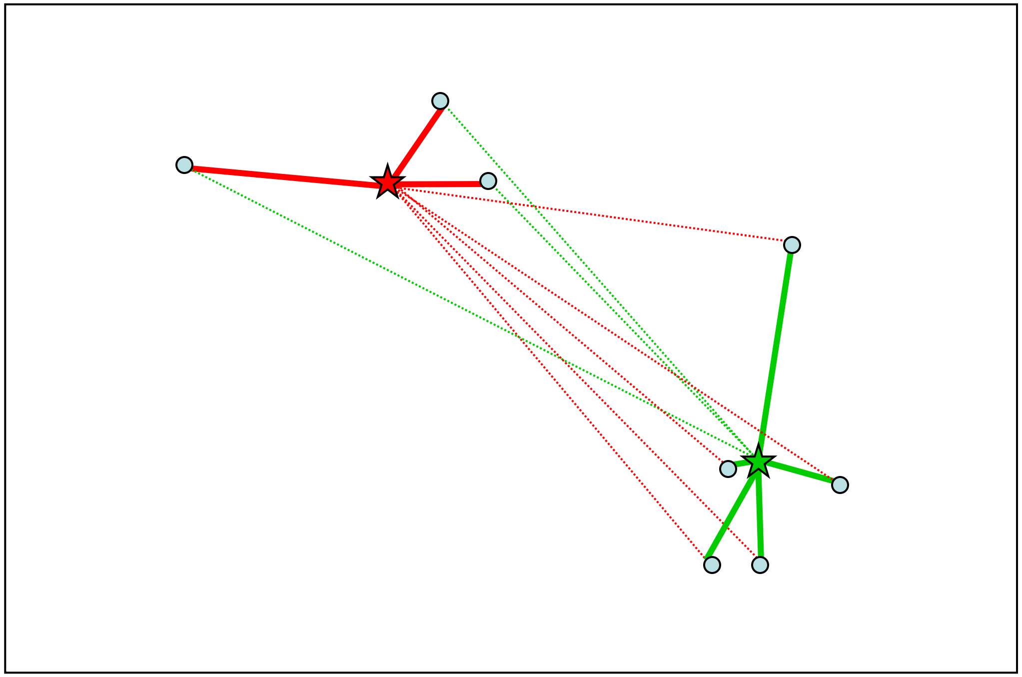
Expectation Maximzation: Idea

- Data generated from mixture of Gaussians



- Latent variables: Correspondence between Data Items and Gaussians

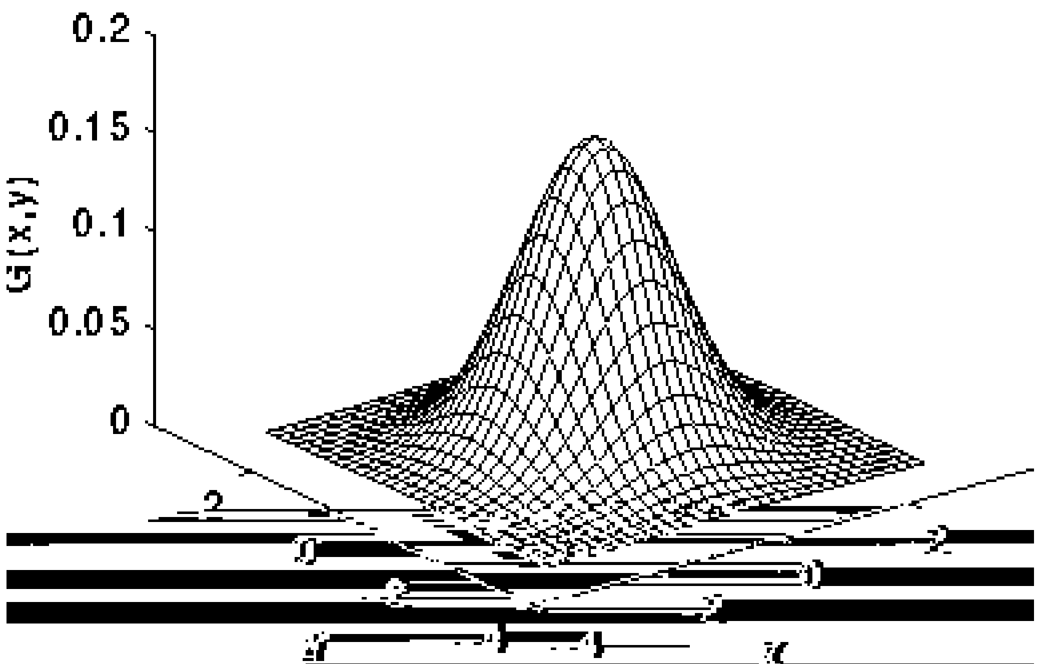
Generalized K-Means (EM)



Gaussians

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

$$p(x) = (2\pi)^{-\frac{N}{2}} \left|\Sigma\right|^{-1} \exp\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right\}$$



ML Fitting Gaussians

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

$$p(x) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-1} \exp\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right\}$$

$$\mu = \frac{1}{M} \sum_i x_i$$

$$\Sigma = \frac{1}{M} \sum_i (x_i - \mu)(x_i - \mu)^T$$

Learning a Gaussian Mixture

(with known covariance)

E-Step

$$E[z_{ij}] \quad \frac{p(x \quad x_i \mid \quad {}_j)}{{}^{k}_{n\ 1}\ p(x \quad x_i \mid \quad {}_n)} \quad \frac{e^{-\frac{1}{2\ ^2}(x_i \quad {}_j)^2}}{{}^{k}_{n\ 1}\ e^{-\frac{1}{2\ ^2}(x_i \quad {}_n)^2}}$$

M-Step

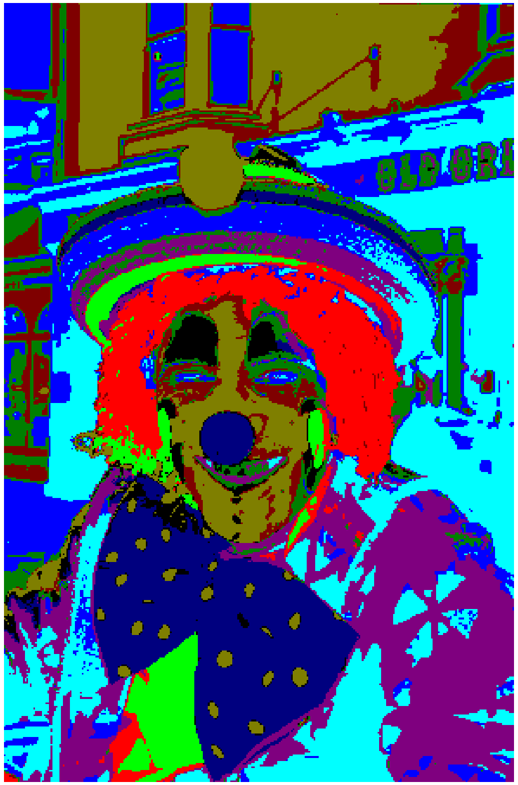
$$\mu_j \leftarrow \frac{1}{n_j}\sum_{i=1}^{M} E[z_{ij}]\, x_i \qquad n_j \leftarrow \sum_{i=1}^{M} E[z_{ij}]$$

$$\Sigma_j \leftarrow \frac{1}{n_j}\sum_{i=1}^{M} E[z_{ij}] \left(x_i - \mu_j\right)\left(x_i - \mu_j\right)^T$$

Expectation Maximization

- Converges!
- Proof [Neal/Hinton, McLachlan/Krishnan]:
  - E/M step does not decrease data likelihood
  - Converges at local minimum or saddle point
- But subject to local minima

EM Clustering: Results



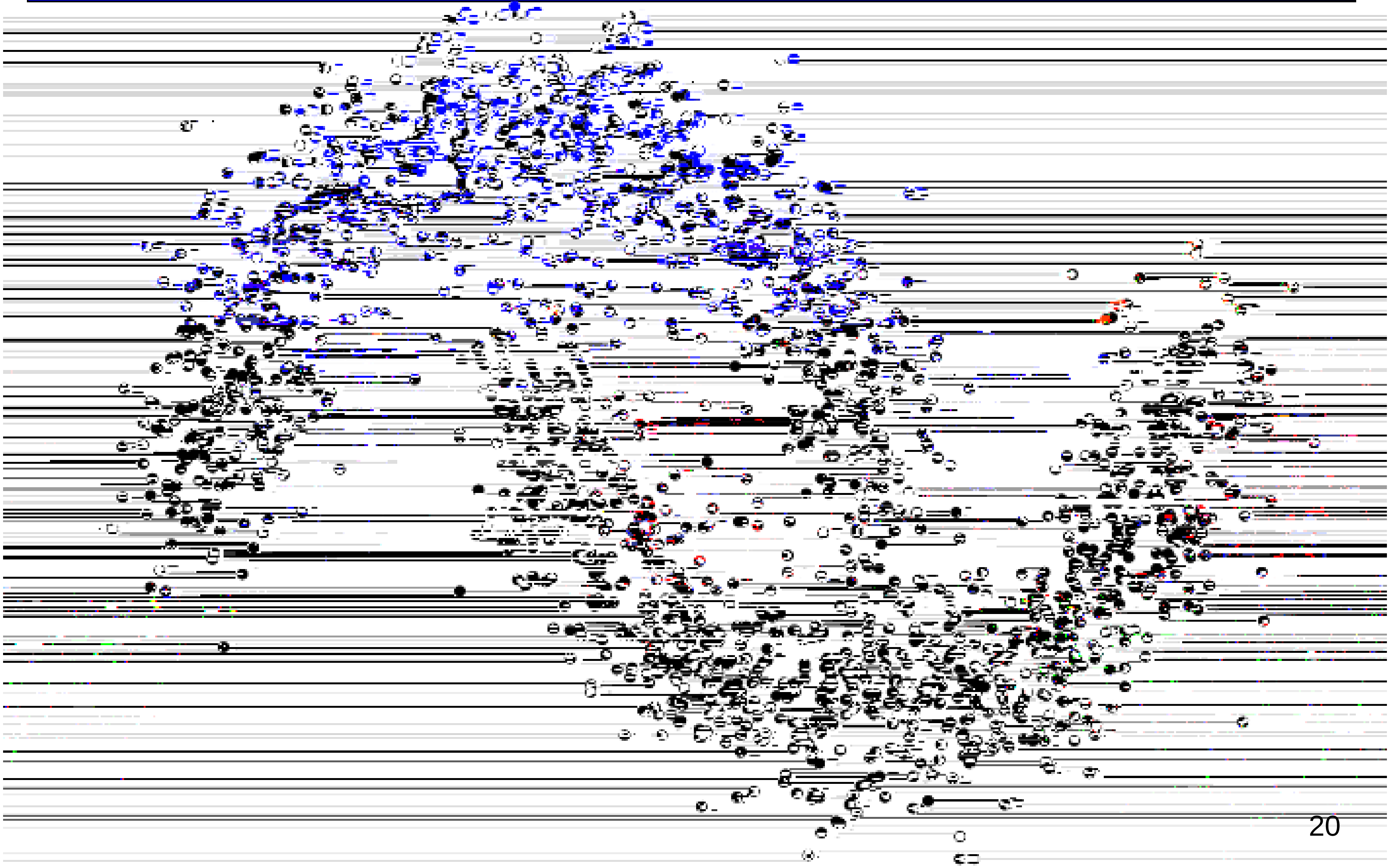
http://www.ece.neu.edu/groups/rpl/kmeans/

Practical EM

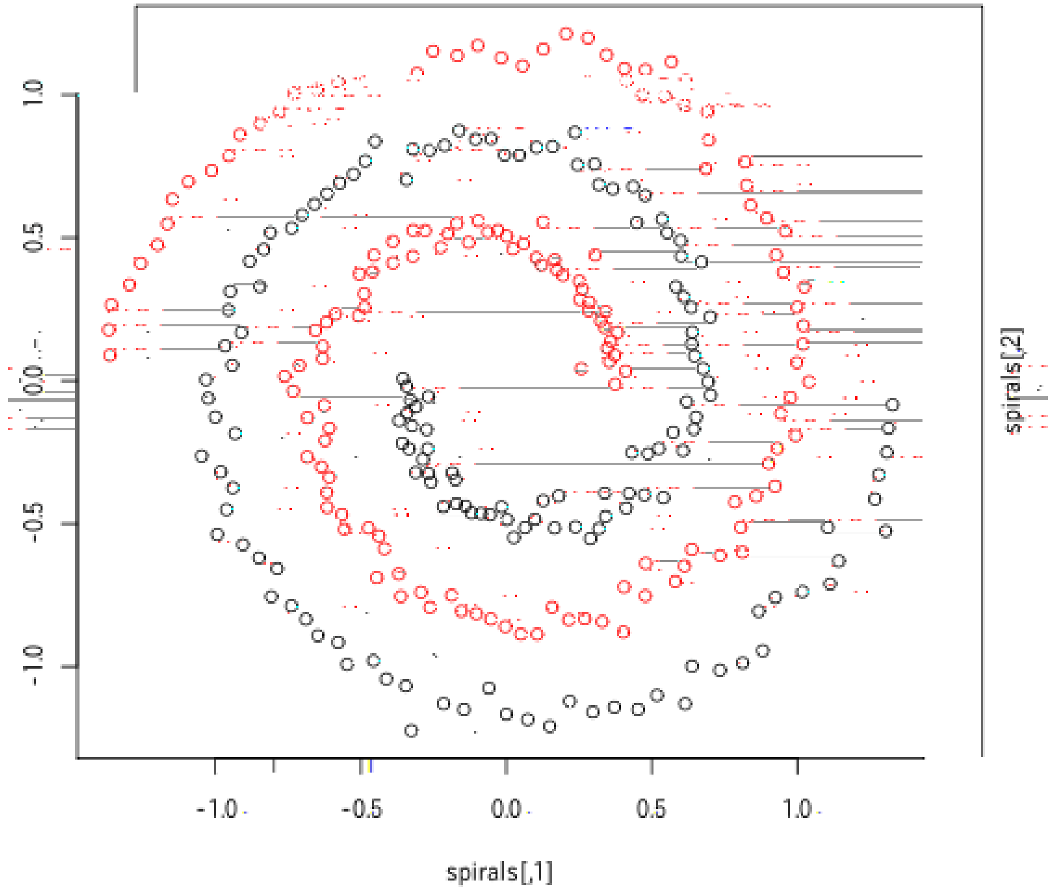
- Number of Clusters unknown
- Suffers (badly) from local minima
- Algorithm:
  - Start new cluster center if many points "unexplained"
  - Kill cluster center that doesn't contribute
  - (Use AIC/BIC criterion for all this, if you want to be formal)

Spectral Clustering

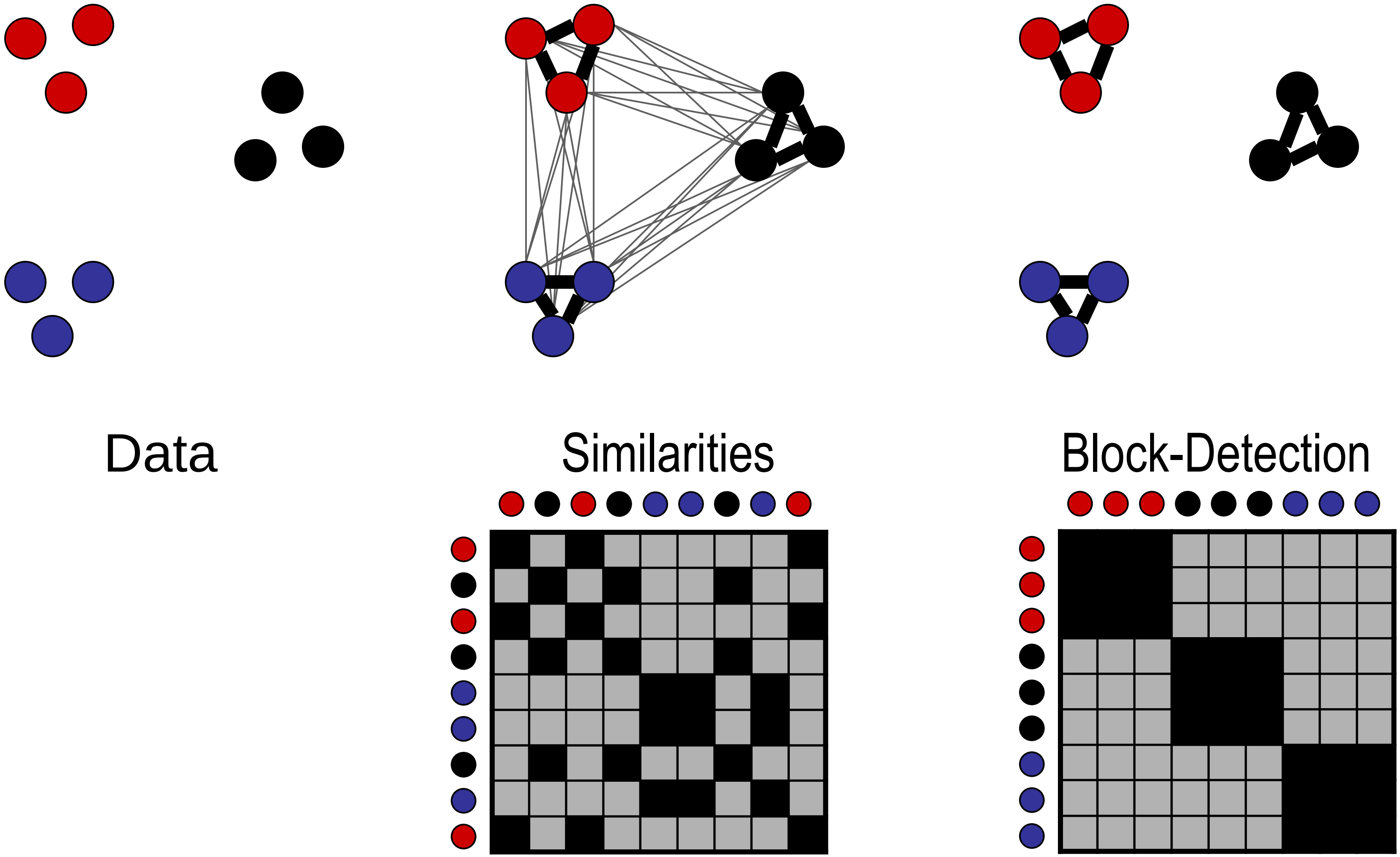
Spectral Clustering



The Two Spiral Problem



Spectral Clustering: Overview



* Slides from Dan Klein, Sep Kamvar, Chris Manning, Natural Language Group Stanford University

Eigenvectors and Blocks

- Block matrices have block eigenvectors:

$$\lambda_1 = 2 \quad \lambda_2 = 2 \quad \lambda_3 = 0 \quad \lambda_4 = 0$$

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \xrightarrow{\text{eigensolver}} \begin{bmatrix} .71 \\ .71 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ .71 \\ .71 \end{bmatrix}$$

- Near-block matrices have near-block eigenvectors: [Ng *et al.*, NIPS 02]

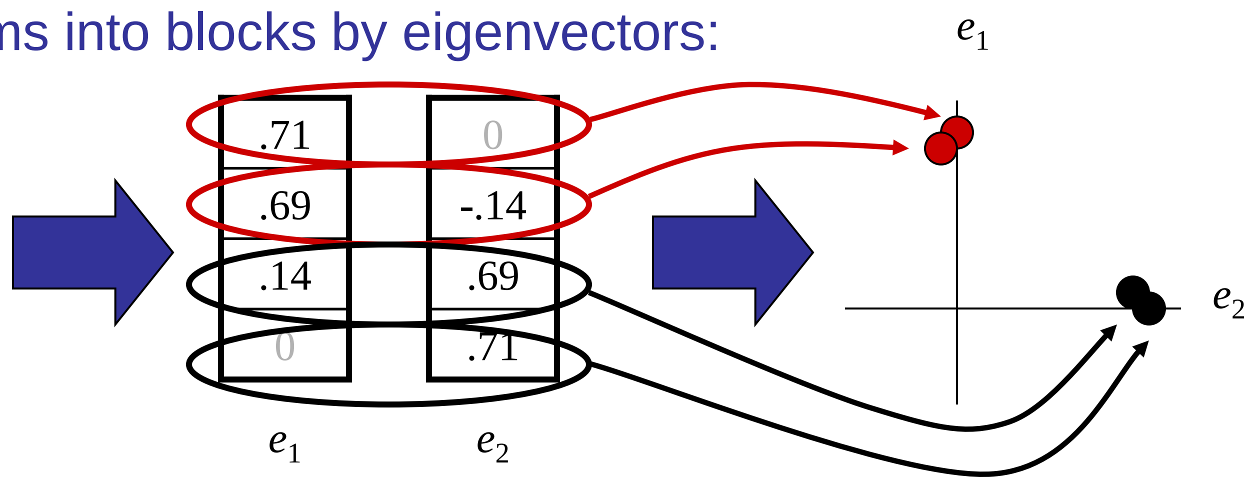
$$\lambda_1 = 2.02 \quad \lambda_2 = 2.02 \quad \lambda_3 = -0.02 \quad \lambda_4 = -0.02$$

$$\begin{bmatrix} 1 & 1 & .2 & 0 \\ 1 & 1 & 0 & -.2 \\ .2 & 0 & 1 & 1 \\ 0 & -.2 & 1 & 1 \end{bmatrix} \xrightarrow{\text{eigensolver}} \begin{bmatrix} .71 \\ .69 \\ .14 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ -.14 \\ .69 \\ .71 \end{bmatrix}$$

Spectral Space

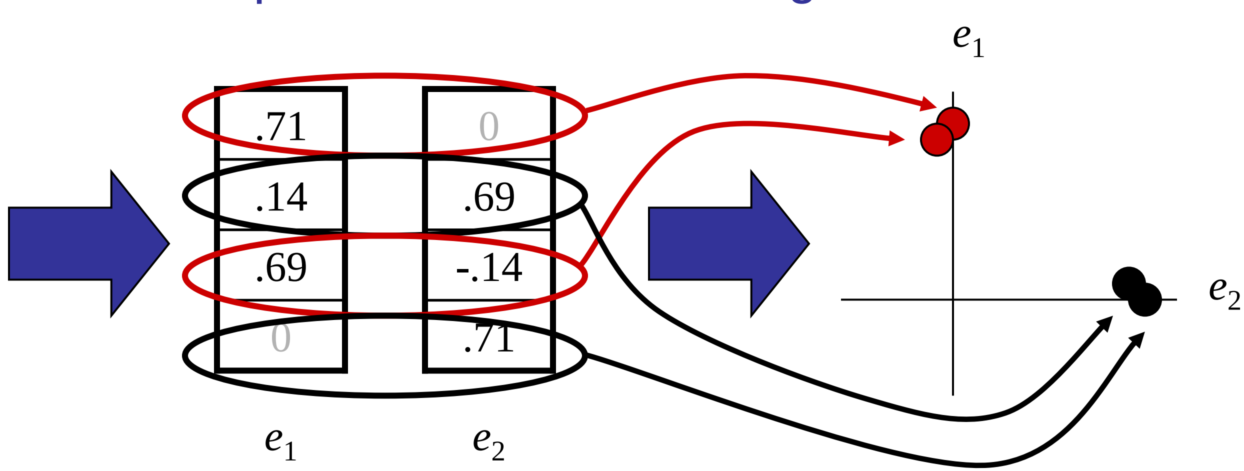
- Can put items into blocks by eigenvectors:

1	1	.2	0
1	1	0	-.2
.2	0	1	1
0	-.2	1	1



- Resulting clusters independent of row ordering:

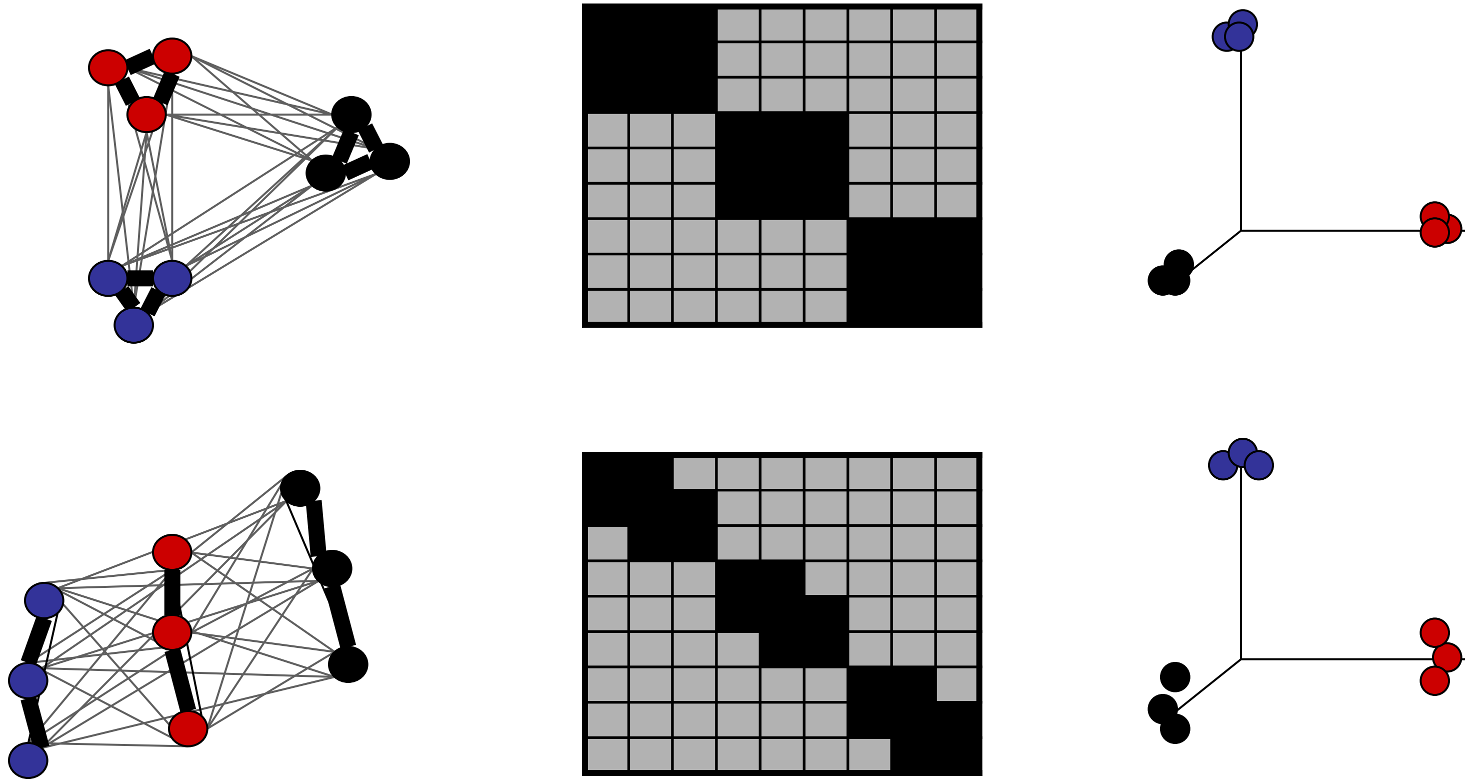
1	.2	1	0
.2	1	0	1
1	0	1	-.2
0	1	-.2	1



* Slides from Dan Klein, Sep Kamvar, Chris Manning, Natural Language Group Stanford University

The Spectral Advantage

- The key advantage of spectral clustering is the spectral space representation:



* Slides from Dan Klein, Sep Kamvar, Chris Manning, Natural Language Group Stanford University

Measuring Affinity

Intensity

$$, \quad \exp \quad {}^{1}\!/_{2} \quad {}^{2} \quad \| \qquad \|^{2}$$

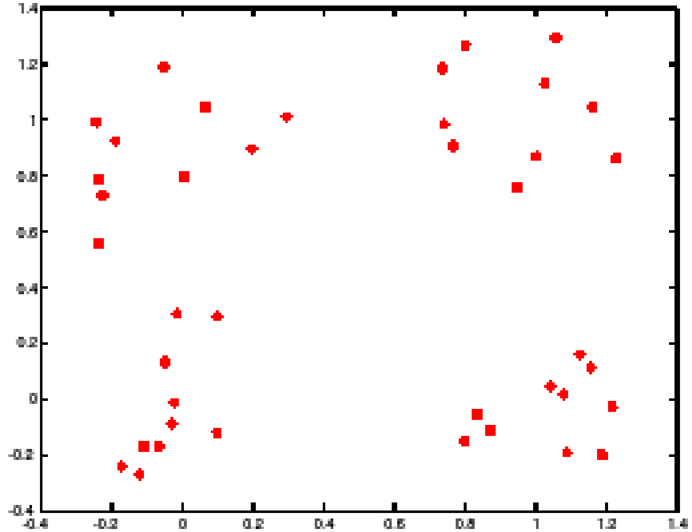
Distance

$$, \quad \exp \quad {}^{1}\!/_{2} \quad {}^{2} \quad \| \qquad \|^{2}$$

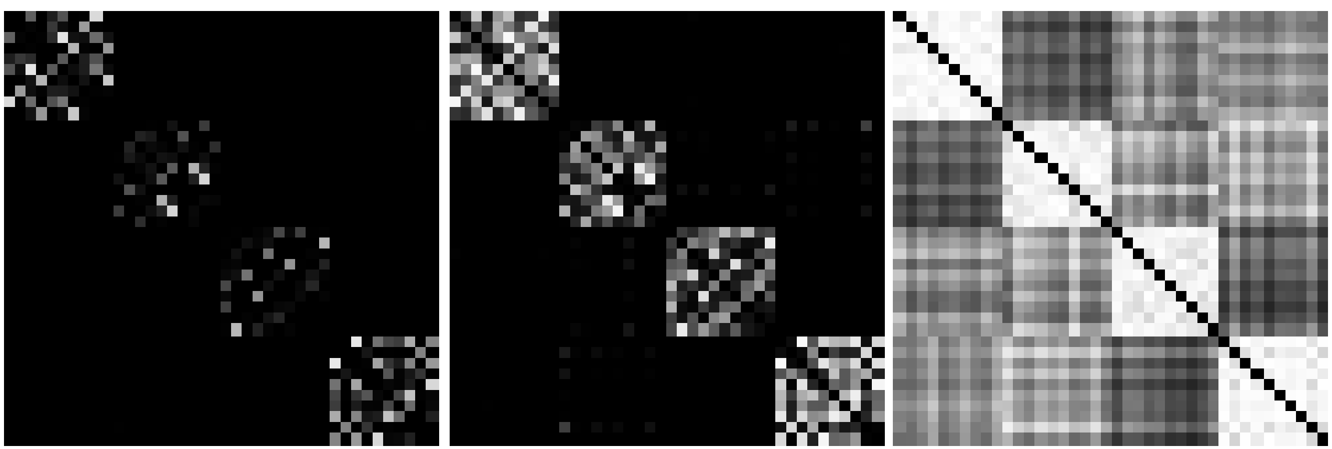
Texture

$$\square, \quad \exp \quad {}^{1}\!/_{2} \quad {}^{2} \quad \| \qquad \|^{2}$$

Scale affects affinity



σ=.2



σ=.1 σ=.2 σ=1

* From Marc Pollefeys COMP 256 2003

Scale affects affinity

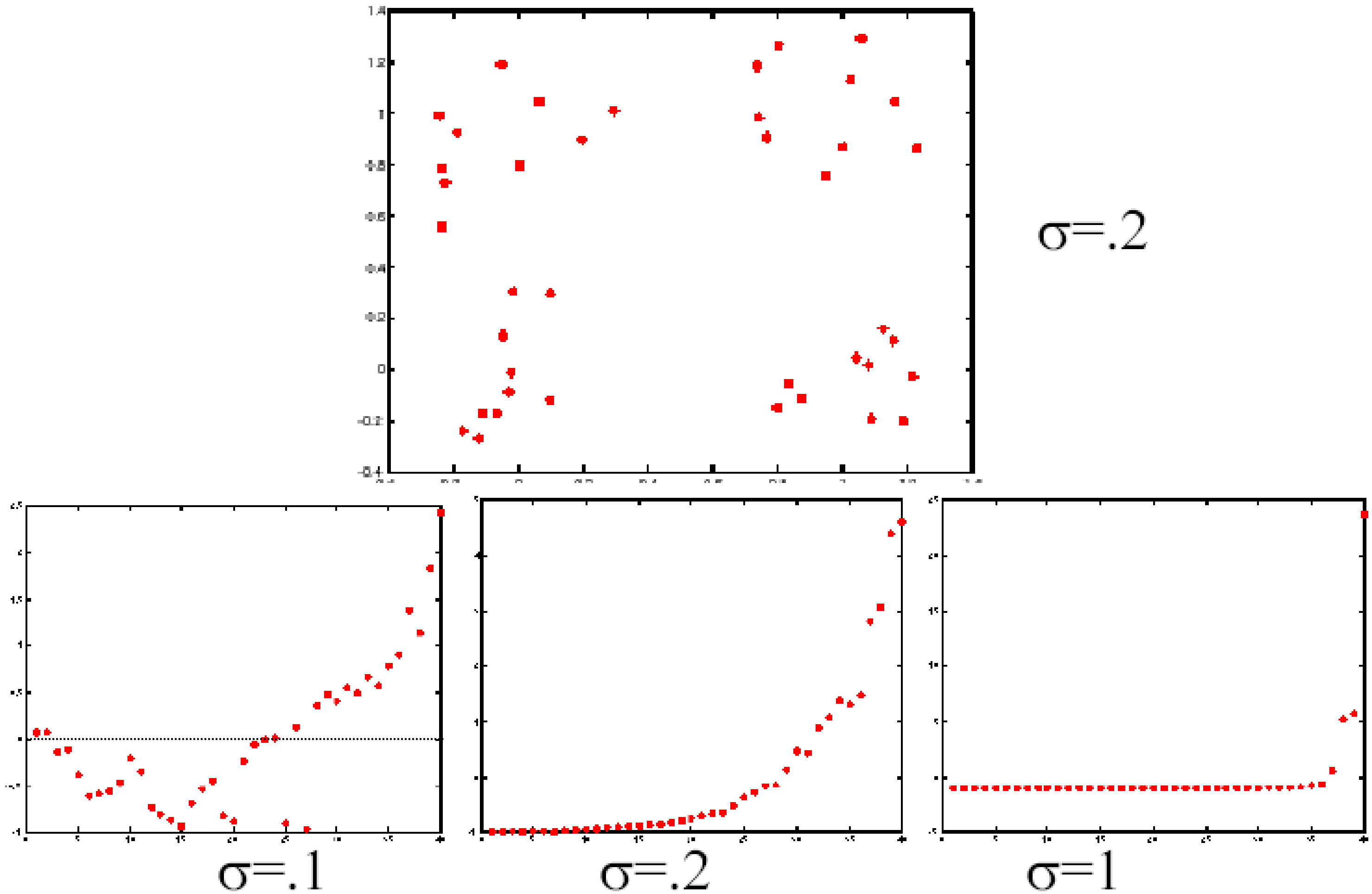
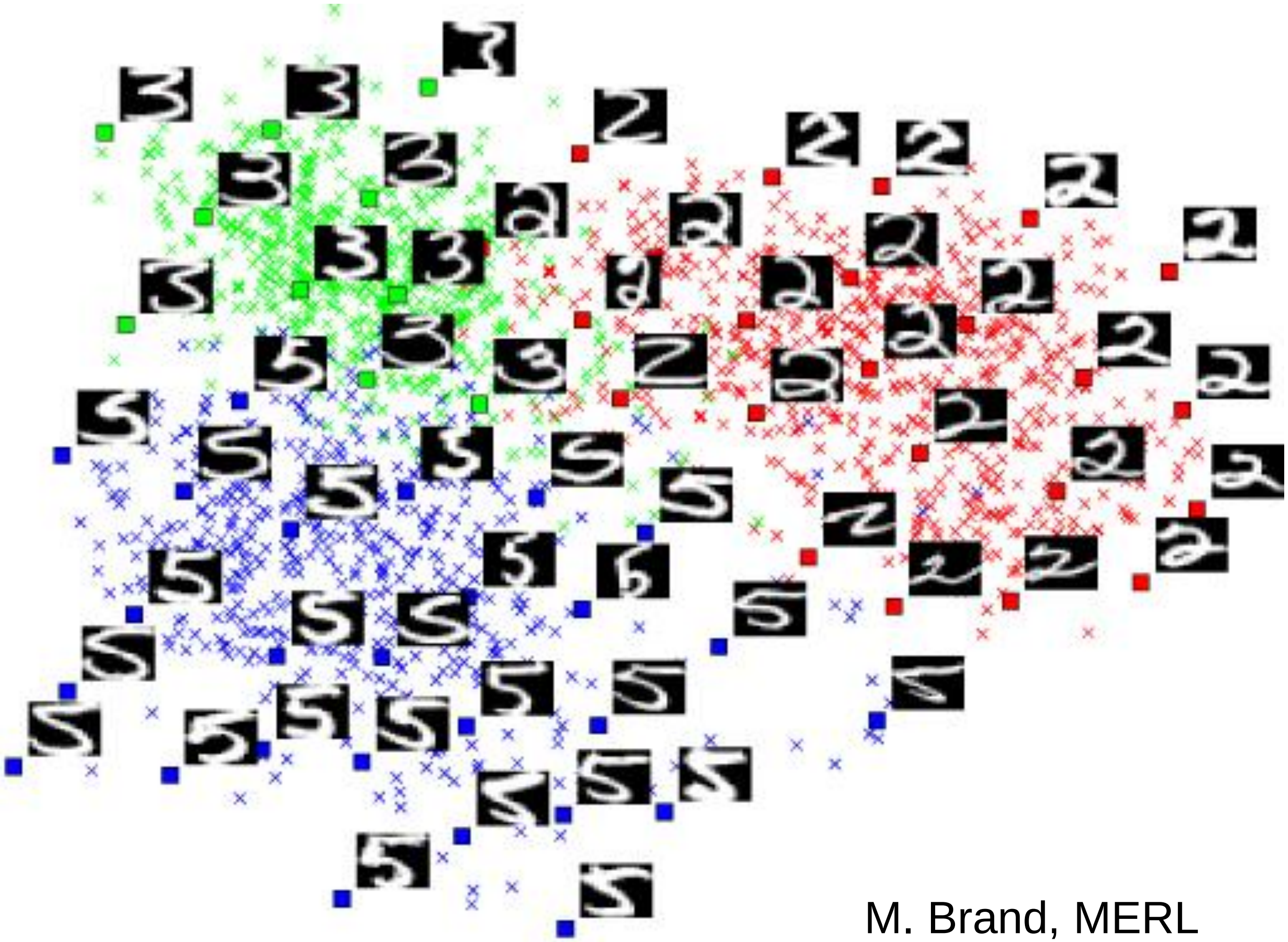


FIGURE 15.21: The number of clusters is reflected in the eigenvalues of the affinity matrix.

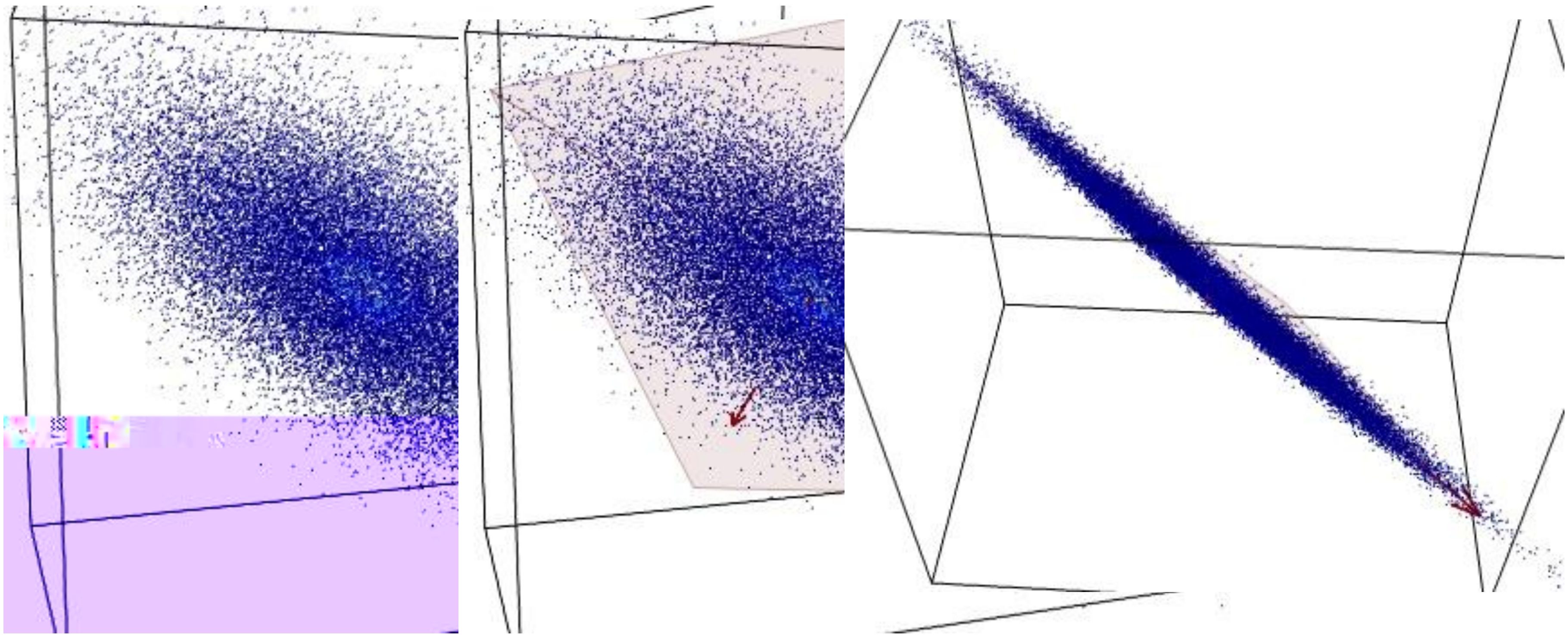
Dimensionality Reduction

The Space of Digits (in 2D)

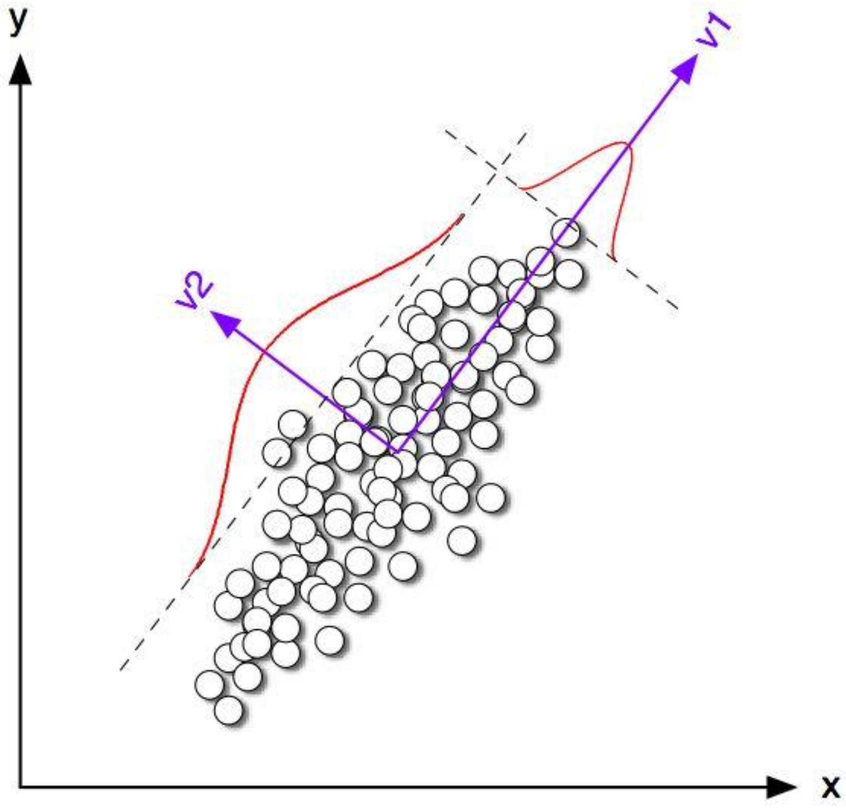


M. Brand, MERL

Dimensionality Reduction with PCA

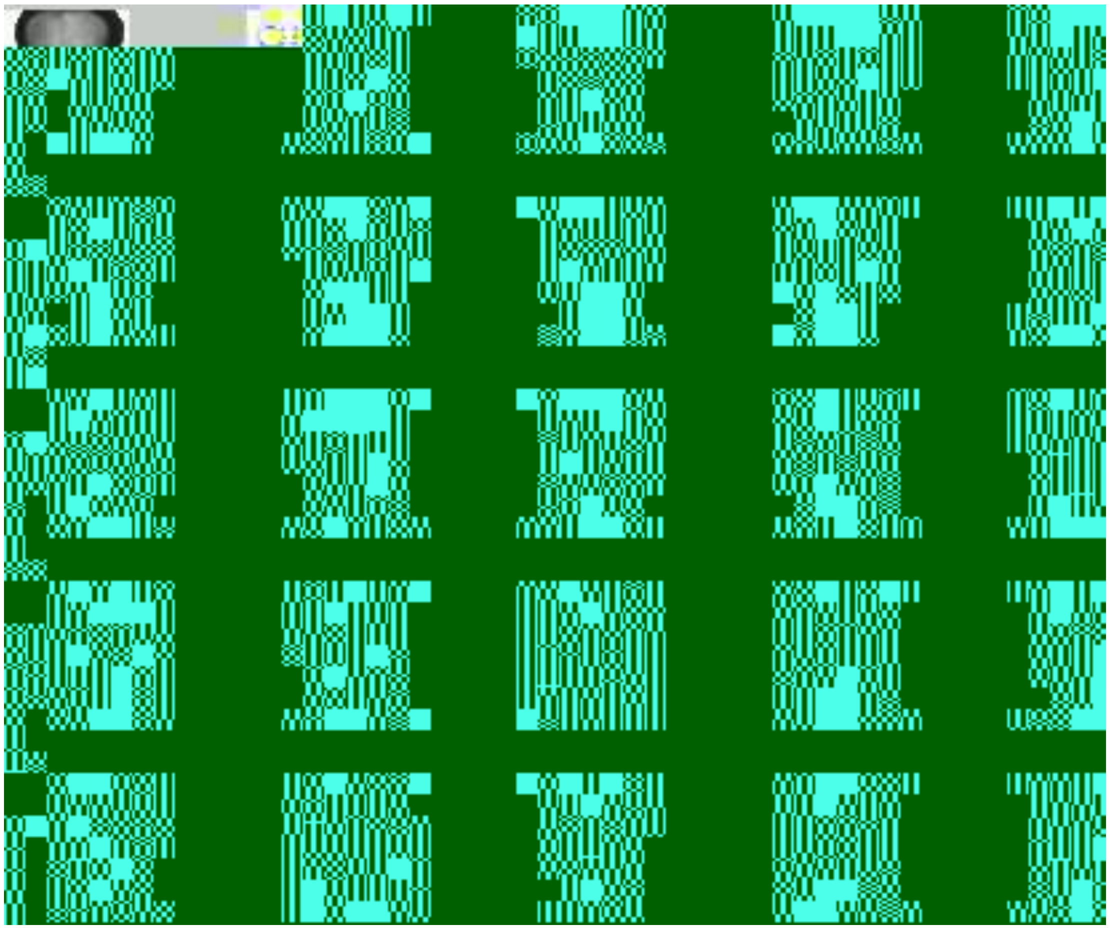


Linear: Principal Components



- Fit multivariate Gaussian
- Compute eigenvectors of Covariance
- Project onto eigenvectors with largest eigenvalues

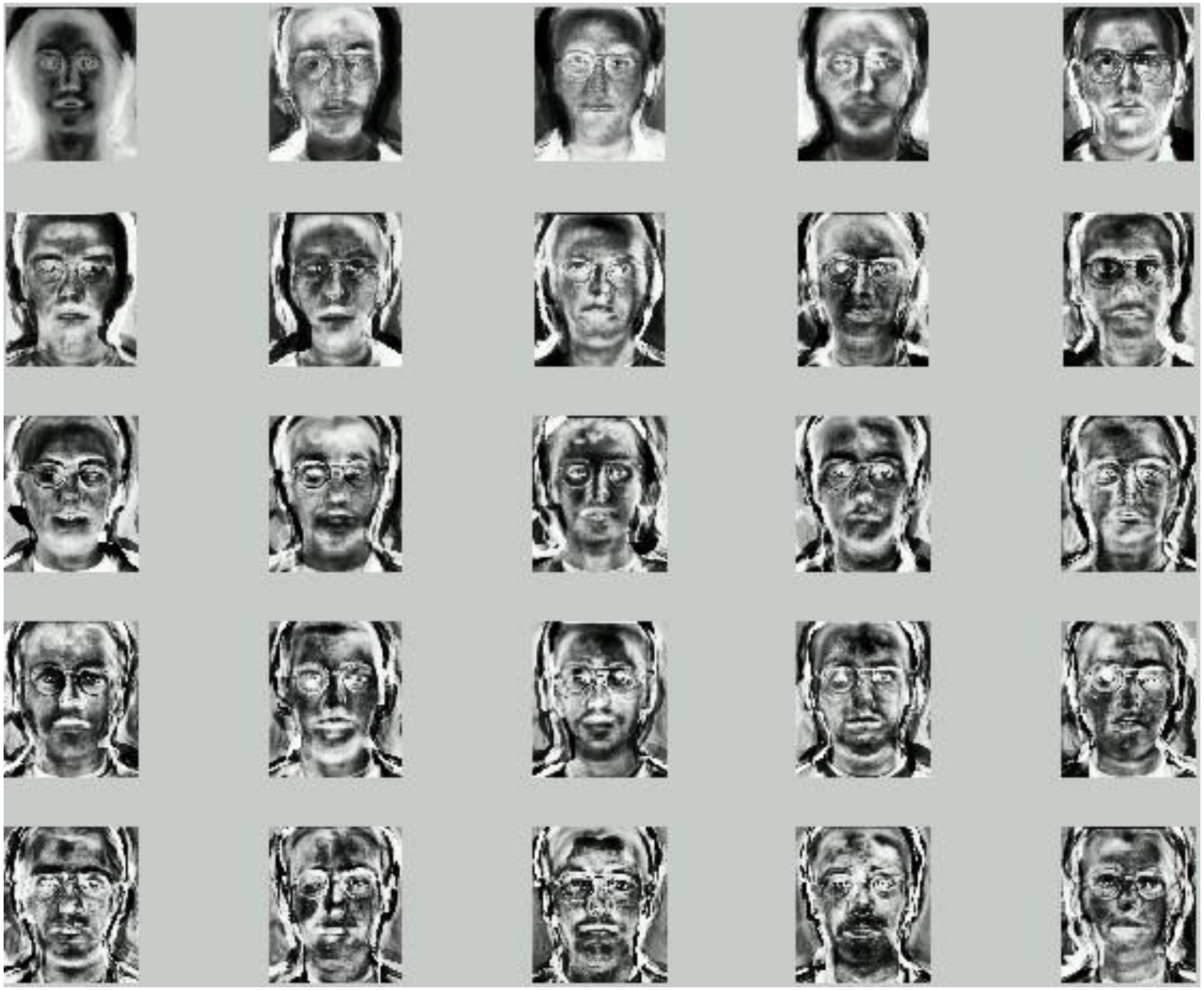
Other examples of unsupervised learning



Mean face (after alignment)

Slide credit: Santiago Serrano

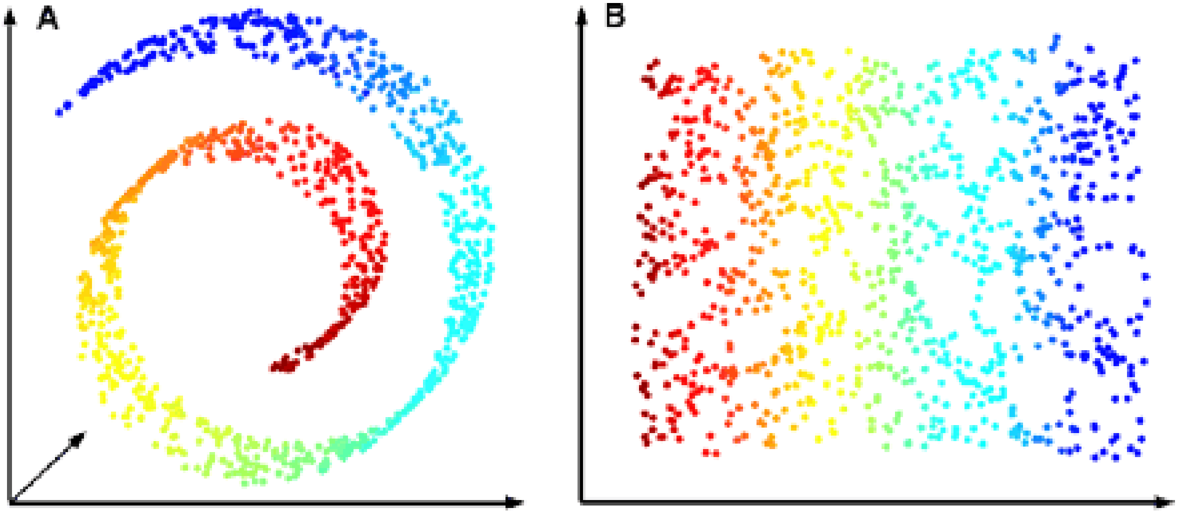
Eigenfaces



Slide credit: Santiago Serrano

Non-Linear Techniques

- Isomap
- Local Linear Embedding



Isomap, Science, M. Balasubmaranian and E. Schwartz

Scape (Drago Anguelov et al)

SCAPE: Shape Completion and Animation of People