Part 1: PCA & MDS

COS 323

Last Time

- How do we solve least-squares…
  - without incurring condition-squaring effect of normal equations ($A^TAx = A^Tb$)
  - when A is singular, "fat", or otherwise poorly-specified?
- QR Factorization
  - Householder method
- Singular Value Decomposition
- Total least squares

Dimensionality Reduction

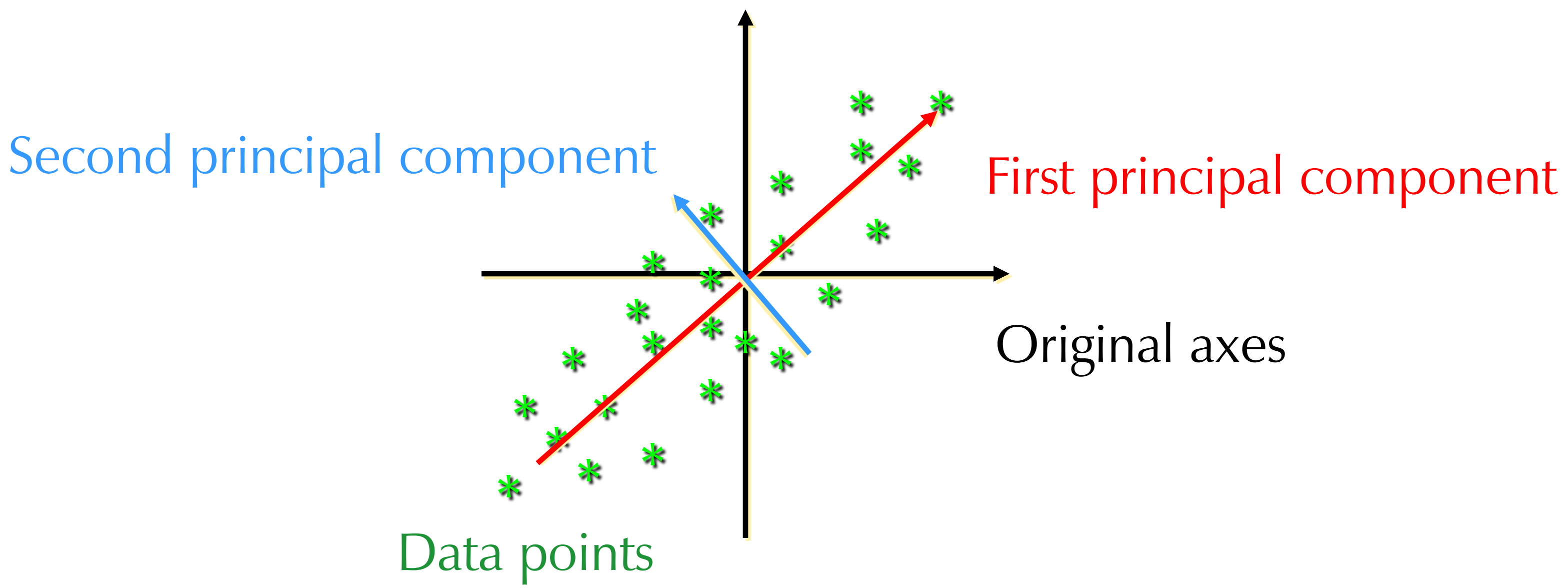
- Map points in high-dimensional space to lower number of dimensions
- Preserve structure: pairwise distances, etc.
- Useful for further processing:
  - Less computation, fewer parameters
  - Easier to understand, visualize

SVD for rank-*k* approximation

- **A** is m×n matrix of rank > k
- Suppose you want to find best rank-*k* approximation to **A**
- Take SVD: $\mathbf{A} = \mathbf{UWV}^T$
- Set all but the largest *k* singular values of **W** to 0
- Can form compact representation by eliminating columns of **U** and **V** corresponding to zeroed w_i

Principal Components Analysis (PCA)

- Approximating a high-dimensional data set with a lower-dimensional linear subspace
- Also converts possibly-correlated attributes into uncorrelated attributes



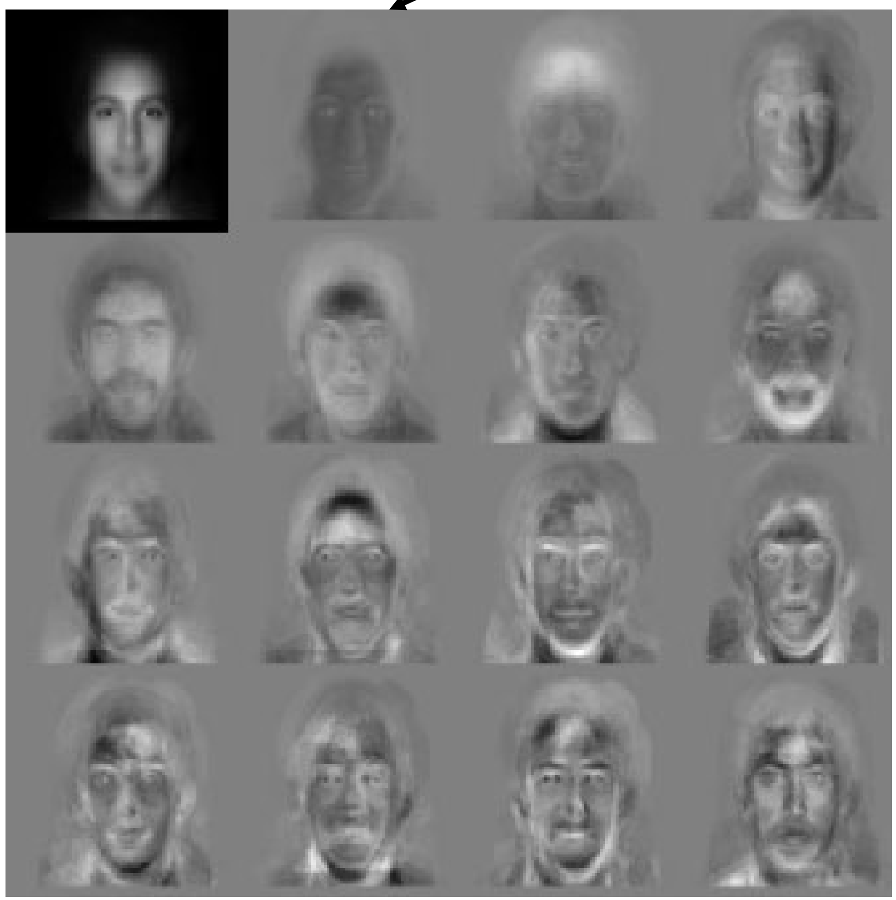
SVD and PCA

- Data matrix with points/examples as rows
- Center data by subtracting mean (“whitening”)
- Compute SVD
- Columns of $\mathbf{V}_k$ are principal components
- Value of w_i gives importance of each component

PCA on Faces: "Eigenfaces"

Average face

First principal component



Other components

For all except average,
"gray" = 0,
"white" > 0,
"black" < 0

Uses of PCA

- Compression: each new image can be approximated by projection onto first few principal components
- Recognition: for a new image, project onto first few principal components, match feature vectors
- Generation: Adjust contributions of a few principal components to generate new plausible data points

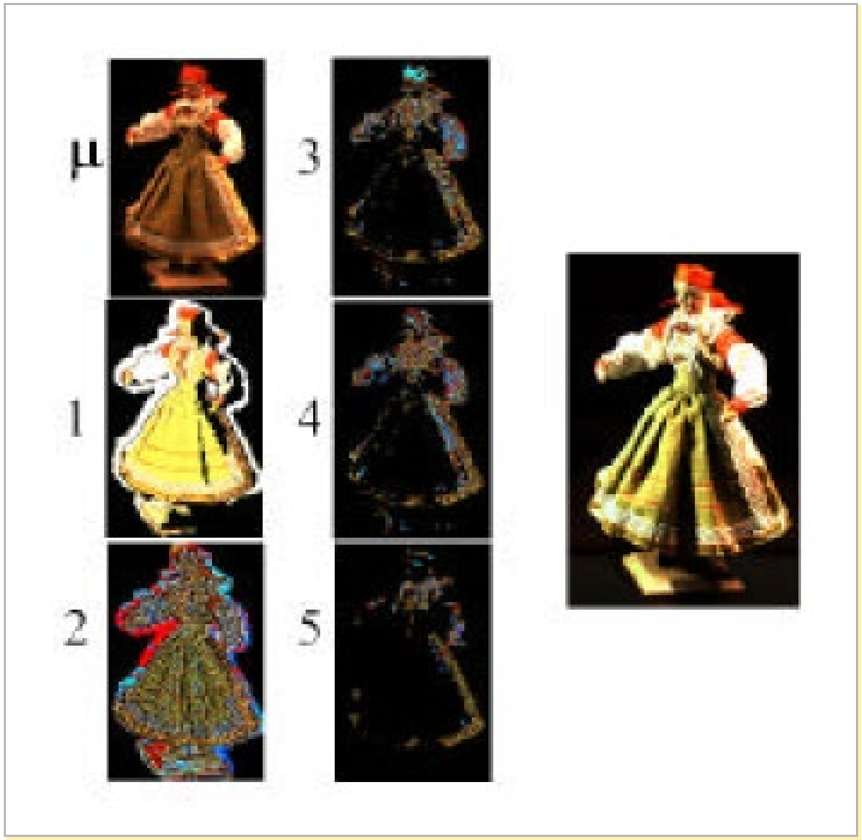
PCA for Relighting

- Images under different illumination

[Matusik & McMillan]

PCA for Relighting

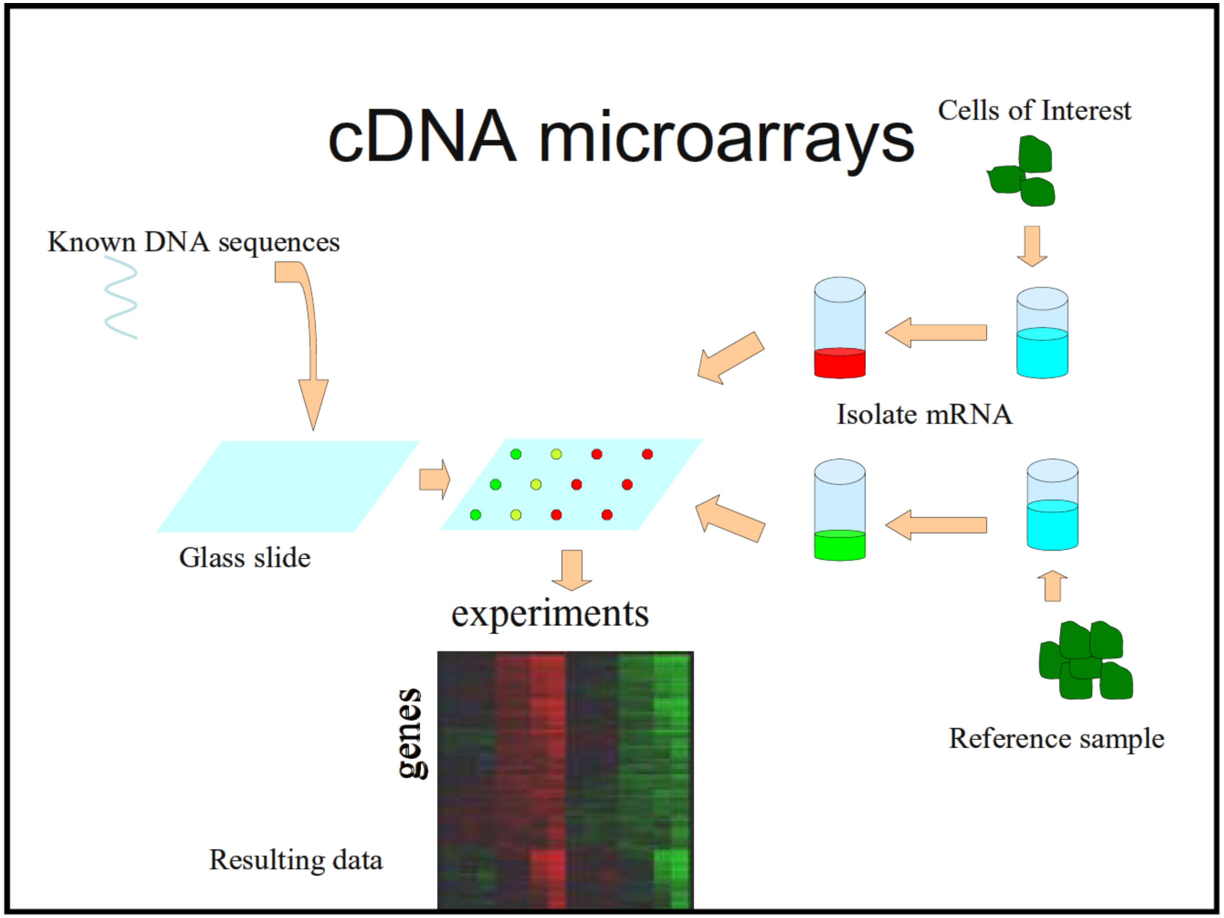
- Images under different illumination
- Most variation captured by first 5 principal components – can re-illuminate by combining only a few images



[Matusik & McMillan]

PCA for DNA Microarrays

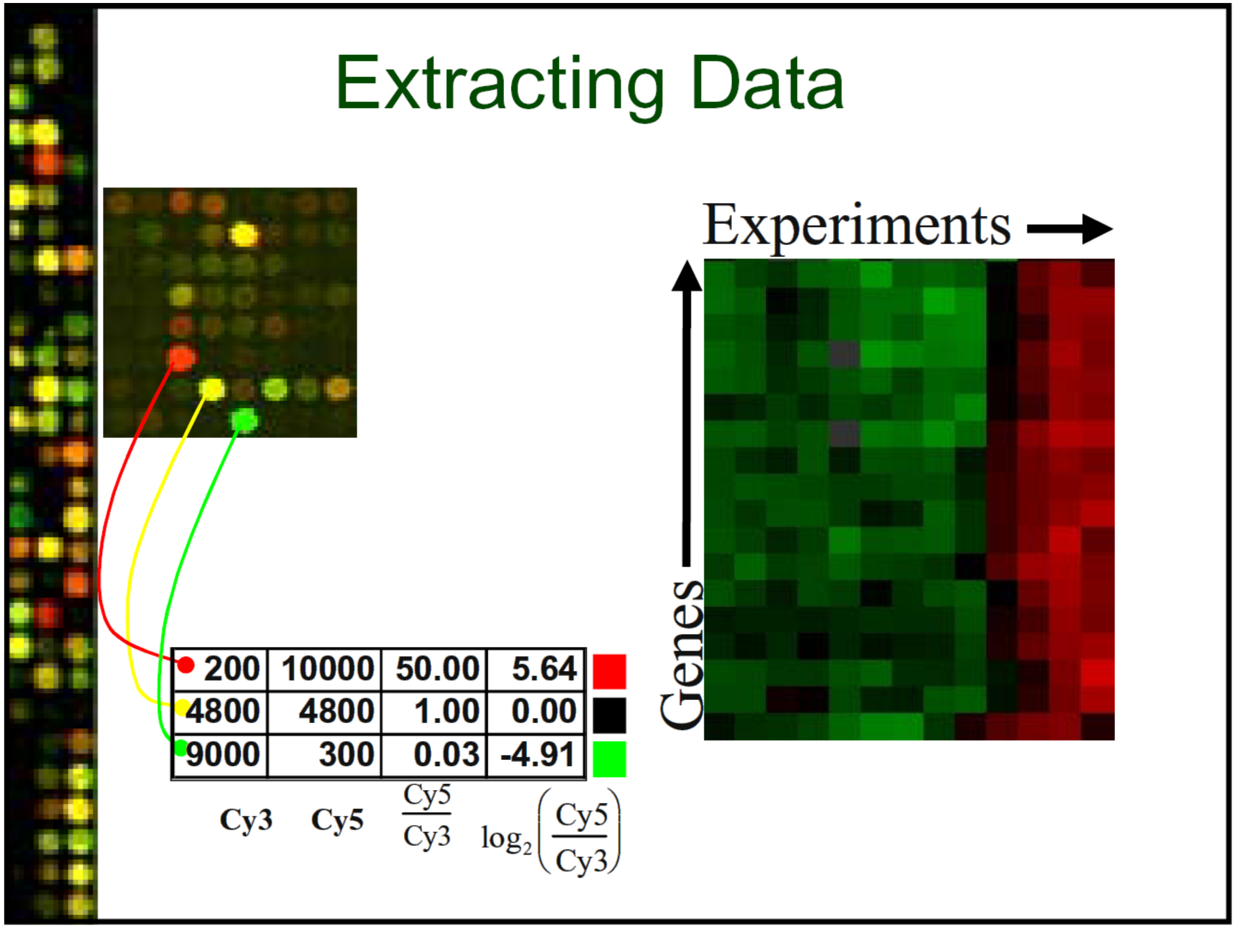
- Measure gene activation under different conditions



[Troyanskaya]

PCA for DNA Microarrays

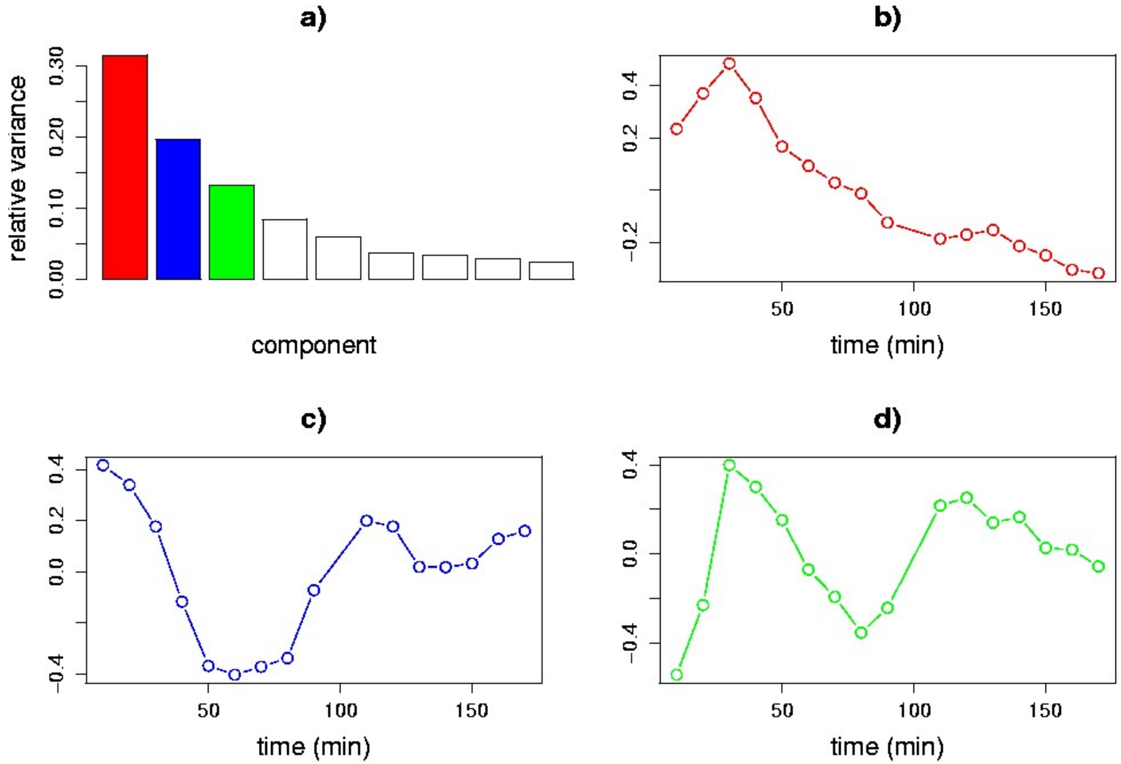
- Measure gene activation under different conditions



[Troyanskaya]

PCA for DNA Microarrays

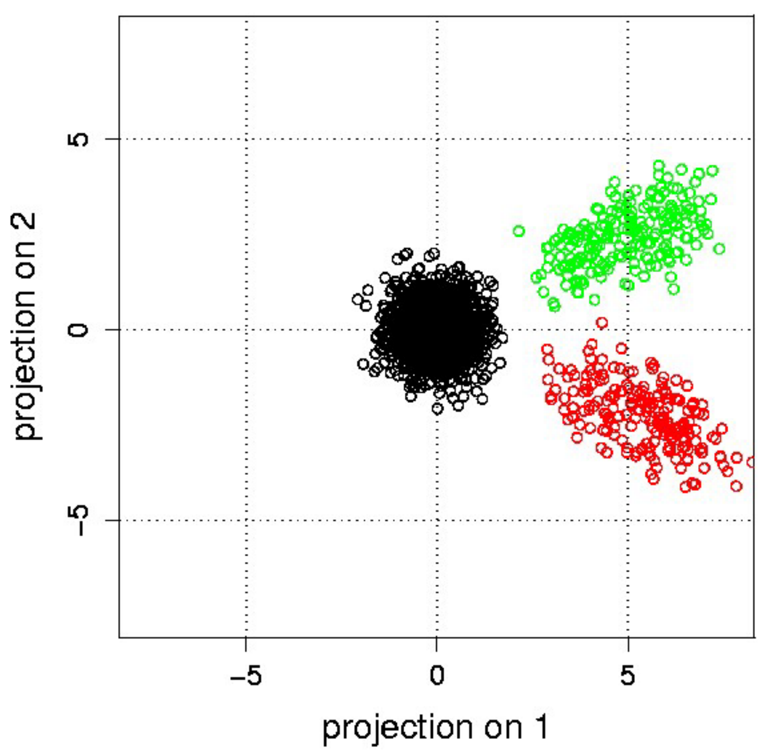
- PCA shows patterns of correlated activation
  - Genes with same pattern might have similar function



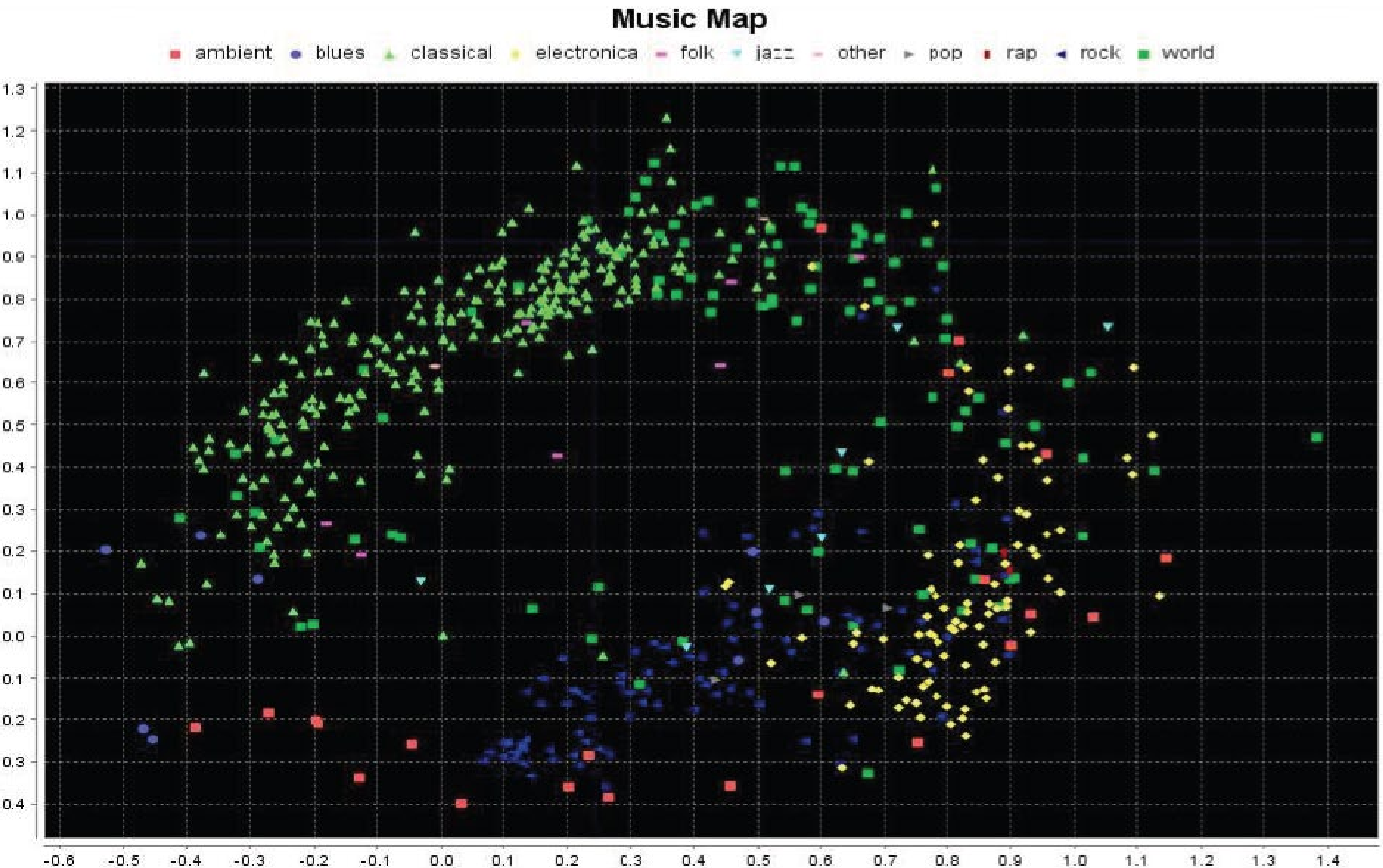
[Wall et al.]

PCA for DNA Microarrays

- PCA shows patterns of correlated activation
  - Genes with same pattern might have similar function



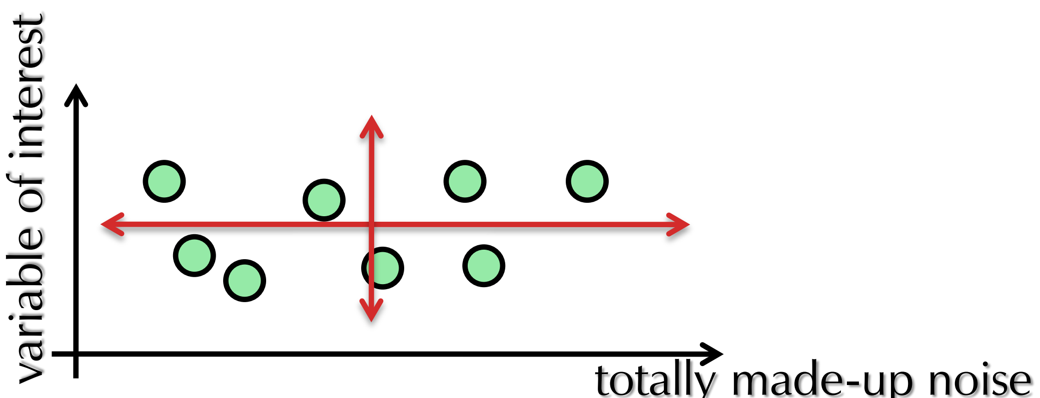
[Wall et al.]



Music Map
ambient
blues
classical
electronica
folk
jazz
other
pop
rap
rock
world
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.6
-0.5
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4

Practical Considerations for PCA

- Sensitive to scale of each attribute (column)
  - In practice, may scale each attribute to have unit variance
- Sensitive to noisy attributes
  - Just because a dimension is highly weighted by PCA doesn't mean it's relevant, informative, etc.



Multidimensional Scaling

Multidimensional Scaling

- In some experiments, can only measure similarity or dissimilarity
  - e.g., is response to stimuli similar or different?
  - Frequent in psychophysical experiments, preference surveys, etc.
- Want to recover absolute positions in k-dimensional space

Multidimensional Scaling

- Example: given pairwise distances between cities

	Atl	Chi	Den	Hou	LA	Mia	NYC	SF	Sea	DC
Atlanta	0									
Chicago	587	0								
Denver	1212	920	0							
Houston	701	940	879	0						
LA	1936	1745	831	1374	0					
Miami	604	1188	1726	968	2339	0				
NYC	748	713	1631	1420	2451	1092	0			
SF	2139	1858	949	1645	347	2594	2571	0		
Seattle	2182	1737	1021	1891	959	2734	2406	678	0	
DC	543	597	1494	1220	2300	923	205	2442	2329	0

  – Want to recover locations

[Pellacini et al.]

Euclidean MDS

- Formally, let's say we have $n \times n$ matrix D consisting of squared distances $d_{ij} = (x_i - x_j)^2$
- Want to recover $n \times k$ matrix X of positions in k-dimensional space

$$D = \begin{pmatrix} 0 & (x_1 - x_2)^2 & (x_1 - x_3)^2 & \\ (x_1 - x_2)^2 & 0 & (x_2 - x_3)^2 & \\ (x_1 - x_3)^2 & (x_2 - x_3)^2 & 0 & \\ & & & \ddots \end{pmatrix}$$

$$X = \begin{pmatrix} (\cdots x_1 \cdots) \\ (\cdots x_2 \cdots) \\ \vdots \end{pmatrix}$$

Euclidean MDS

- Observe that

$$d_{ij}^2 = (x_i - x_j)^2 = x_i^2 - 2x_i x_j + x_j^2$$

- Strategy: convert matrix D of d_{ij}^2 into matrix B of $x_i x_j$
  - "Centered" distance matrix
  - $B = XX^T$

Euclidean MDS

- Centering:
  - Sum of row *i* of *D* = sum of column *i* of *D* =

$$s_i = \sum_j d_{ij}^2 = \sum_j x_i^2 - 2x_i x_j + x_j^2$$

$$= nx_i^2 - 2x_i \sum_j x_j + \sum_j x_j^2$$

  - Sum of all entries in D =

$$s = \sum_i s_i = 2n \sum_i x_i^2 - 2\left(\sum_i x_i \right)^2$$

Euclidean MDS

- Choose $\Sigma x_i = 0$
  - Solution will have average position at origin

$$s_i = nx_i^2 + \sum_j x_j^2, \quad s = 2n\sum_j x_j^2$$

  - Then,

$$d_{ij}^2 - \tfrac{1}{n}s_i - \tfrac{1}{n}s_j + \tfrac{1}{n^2}s = -2x_i x_j$$

- So, to get *B*:
  - compute row (or column) sums
  - compute sum of sums
  - apply above formula to each entry of *D*
  - Divide by –2

Factoring $B = XX^T$ using SVD

- Now have B, want to factor into XX^T
- If X is $n \times k$, B must have rank k
- Take SVD, set all but top k singular values to 0
  - Eliminate corresponding columns of U and V
  - Have $B' = U'W'V'^T$
  - B' is square and symmetric, so $U' = V'$
  - Take $X = U'$ times square root of W'

Multidimensional Scaling

- Result ($k = 2$):

[Pellacini et al.]

Another application

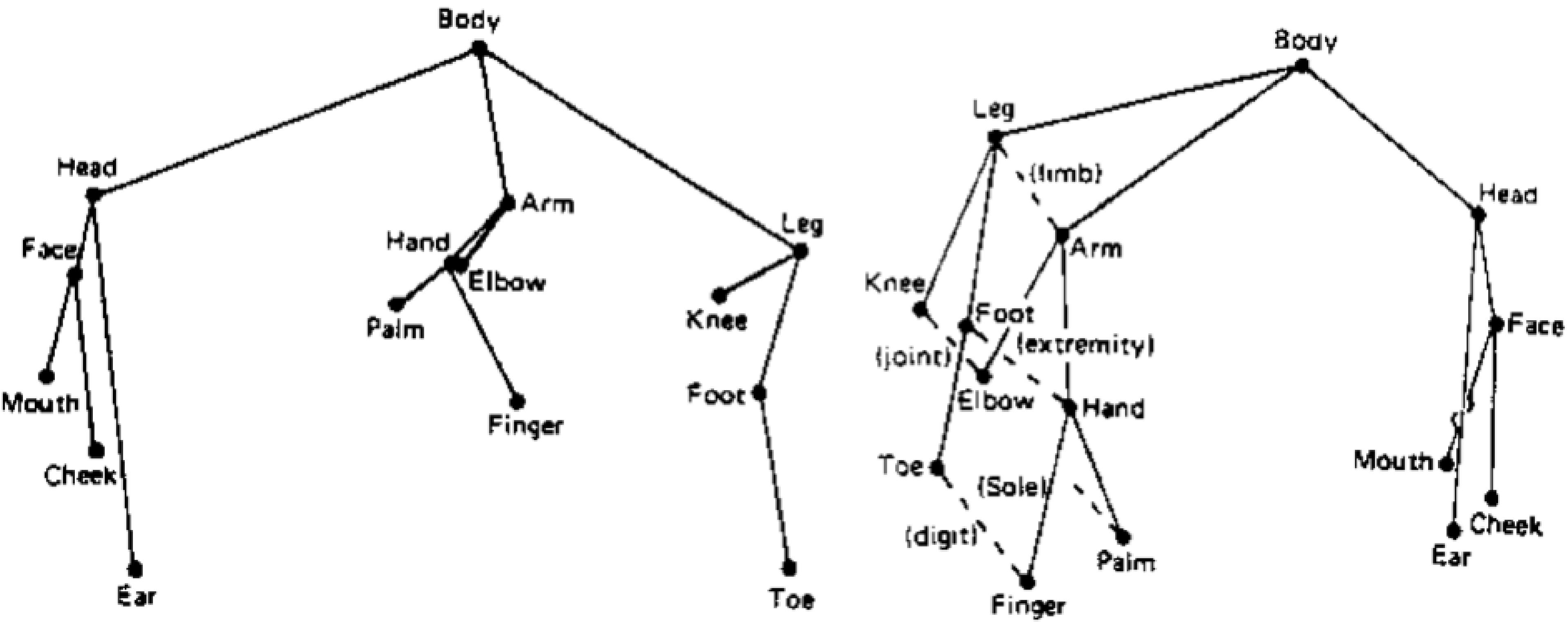
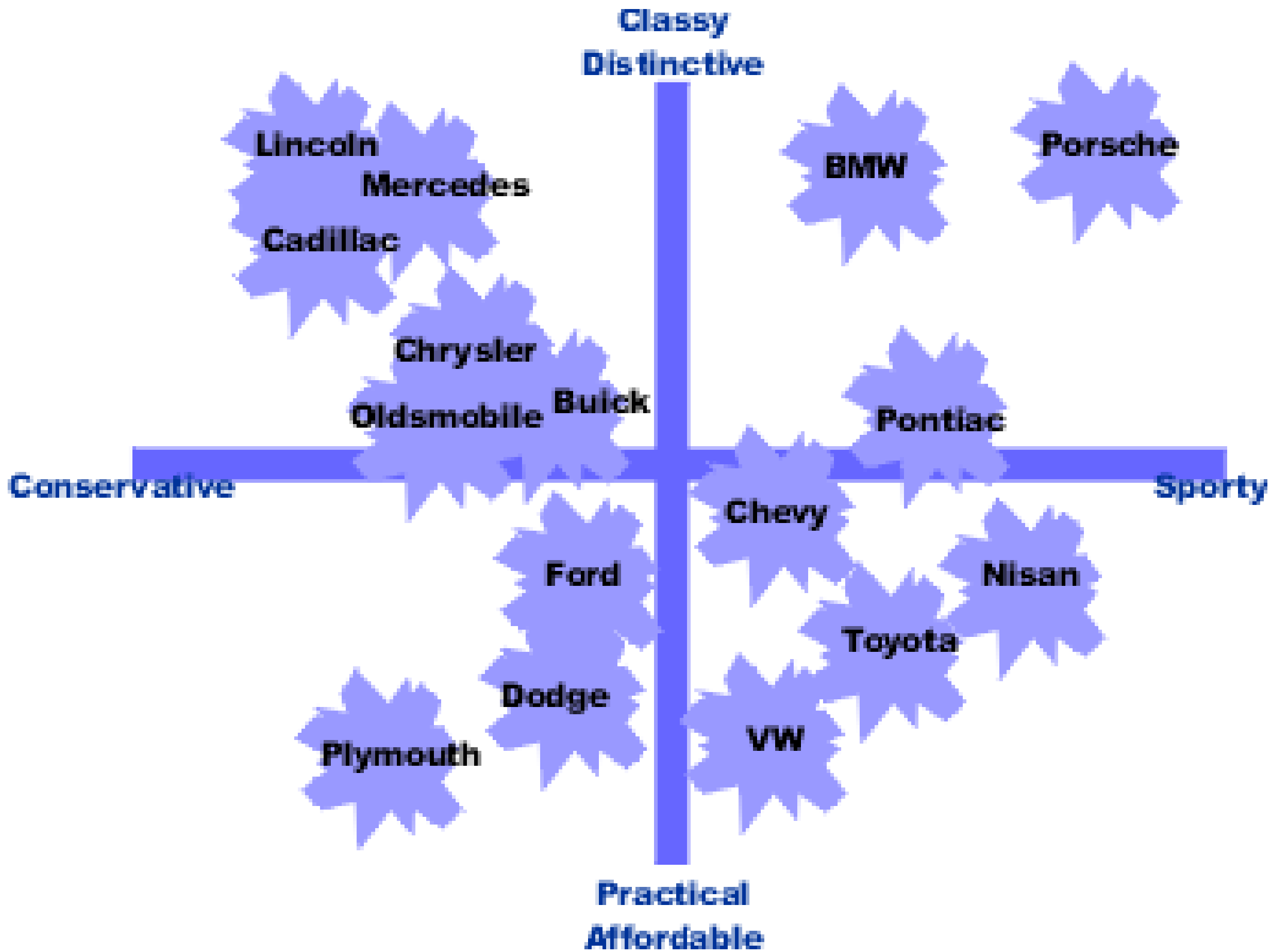


Figure 2 (a) RMDS of children's similarity judgments about 15 body parts; (b) RMDS of adults' similarity judgments about 15 body parts.

From Young 1985 / Jacobowitz 1973

Perceptual Mapping for Marketing



Multidimensional Scaling

- Caveat: actual axes, center not necessarily what you want (can't recover them!)
- This is "classical" or "Euclidean" MDS [Torgerson 52]
  - Distance matrix assumed to be actual Euclidean distance
- More sophisticated versions available
  - "Non-metric MDS": not Euclidean distance, sometimes just *inequalities*
  - Replicated MDS: for multiple data sources (e.g. people)
  - "Weighted MDS": account for observer bias