Part 1: PCA & MDS

COS 323
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 $m \times n$ matrix of rank $> k$
- Suppose you want to find best rank-$k$ approximation to $A$
- Take SVD: $A = U W V^T$
- Set all but the largest $k$ singular values of $W$ to zero
- 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 $V_k$ are principal components
- Value of $w_i$ gives importance of each component
PCA on Faces: “Eigenfaces”

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
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
PCA for DNA Microarrays

- Measure gene activation under different conditions
PCA for DNA Microarrays

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

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

[Wall et al.]
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

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

- Want to recover locations
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

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

$$X = \begin{pmatrix}
(\cdots x_1 \cdots) \\
(\cdots x_2 \cdots) \\
\vdots
\end{pmatrix}$$
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 $D = \text{sum of column } i \text{ of } D =$

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

\[
= nx_i^2 - 2x_i \sum_j x_j + \sum_j x_j^2
\]

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

- 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 ($k = 2$):
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