Pattern Change Discovery between High Dimensional Data Sets

Yi Xu
yxu@cs.binghamton.edu
Computer Science
Binghamton University
What is High Dimensional data?
Low Dimensional Data

People's Height:
-  Napoleon: 5' 6" (1.68m)
- Barack Obama: 6' 1" (1.85m)
- Gordon Brown: 5' 11" (1.80m)
- Dmitry Medvedev: 5' 4" (1.63m)
- Nicolas Sarkozy: 5' 5" (1.65m)
- Silvio Berlusconi: 5' 5" (1.65m)

Magnitude Matters!

1 feature:
- meter

Stock Price:

1 feature:
- price

Gaussian Noise:
- Smart ticks - 2d gaussian noise

2 features: x and y
To view a data sample as a vector

Vector Space Constructed by Ordered Features

<table>
<thead>
<tr>
<th>Text*</th>
<th>Imag**</th>
</tr>
</thead>
<tbody>
<tr>
<td>To</td>
<td>(1,1)</td>
</tr>
<tr>
<td>Be</td>
<td>(1,2)</td>
</tr>
<tr>
<td>Or</td>
<td>(1,3)</td>
</tr>
<tr>
<td>Not</td>
<td>(2,1)</td>
</tr>
<tr>
<td>That</td>
<td>(2,2)</td>
</tr>
<tr>
<td>Is</td>
<td>(2,3)</td>
</tr>
<tr>
<td>A</td>
<td>(3,1)</td>
</tr>
<tr>
<td>Question</td>
<td>(3,2)</td>
</tr>
<tr>
<td></td>
<td>(3,3)</td>
</tr>
</tbody>
</table>

*Hamlet = (2,2,1,1,1,1,1,1,1)ᵀ

**=(0,30,60,210,255,90,180,150,120)ᵀ
Some High Dimensional Feature Space:

- Vocabulary of one month news from New York Times/Politics:
  \[ >8000 \]

- An image data set containing images with 600x800 resolution:
  \[ =480,000 \]

- Algae genome data set acquired using DNA microarray:
  \[ >10^7 \]
Magnitude matters?
Or something else?
Difference between two vectors involves both the magnitude and the direction:

\[ |x_1 - x_2| \]

The Euclidean metric fails to differentiate the length difference from the direction difference.
For high dimensional data, the rotation of a subspace is different from and usually more informative than the magnitude difference.

--- Breaking news – new combination of key words

--- A baby v.s. An adult v.s. A monkey

What measurement is invariant under data’s magnitude change and only characterize the rotation introduced by dimensionality?

- Euclidean Distance
- L-Norms
- Bregman Divergence
How to compute subspace rotation (pattern change) between two vector sets?

\[ \{ \mathbf{x}_i \}_{i=1}^{70} \]

\[ \{ \mathbf{y}_i \}_{i=1}^{50} \]
Principal angles between two subspaces (Golub and Loan)

**Definition 1.** Let $S_1$ and $S_2$ be subspaces in $\mathbb{R}^n$ whose dimensions satisfy

$$p = \text{dim}(S_1) \geq \text{dim}(S_2) = q \geq 1$$

The principal angles $\theta_k \in [0, \pi/2]$, $k = 1, \ldots, q$, between $S_1$ and $S_2$ are defined recursively as

$$\cos(\theta_k) = \max_{u \in S_1, v \in S_2} u^T v = u_k^T v_k$$

when $k = 1$, $\|u_1\| = \|v_1\| = 1$; when $k \geq 2$, $\|u_k\| = \|v_k\| = 1$; $u_k^T u_i = 0$; $v_k^T v_i = 0$ where $i = 1, \ldots, k - 1$. 
Principal angles between two subspaces (cont.)

- A generalization of the angle between two vectors
- A unique set of angles $\{ \cos^{-1} \theta \}_{i=1}^{k}$ defined using orthonormal basis of two subspaces
- Invariant under an isomorphism and thus independent of the magnitude change.
- The largest angles reflect the pattern change between two data sets
- No statistical significance
- Sensitive to noise
Not practical to directly compute the largest principal angles.

What we really want?

Principal angles between subspaces of high likelihood.
Prior information:

- Data

- The number of patterns $k$ that we recognize from the data, usually based on common sense and widely acknowledged facts.

- Label of the data (classification problem)
Pattern Summary via Matrix Factorization

\[ X \approx PS^T \]

\( P \) --- Each column vector represents a summarized pattern from \( X \)

\( S \) --- Each row vector represents the weight on each pattern to restore the corresponding sample.
The subspace $\text{span}(P)$ is where we find principal angles.
Given two data sets $X_1$ and $X_2$

First compute

$$X_1 \approx P_1 S_1^T \quad X_2 \approx P_2 S_2^T$$

Then compute principal angles between $\text{span}(P_1)$ and $\text{span}(P_2)$

Still does not know what is new in subspace $\text{span}(P_2)$.
Given two data sets \( X_1 \) and \( X_2 \), how do we find \( \text{span}(P_2) \), which has the largest principal angles from \( \text{span}(P_1) \)?

Construct hypothesis test on principal angles between \( \text{span}(P_1) \) and \( \text{span}(P_2) \).
Pattern Change Detection Based on Hypothesis Test

- Construct null-hypothesis using principal angles between span(\(P_1\)) and span(\(P_2\)):

\[
H_0 : \left\| \text{diag}\left(\{\cos \theta_i\}_{i=1}^k\right) \right\| = 0
\]

- \(H_0\) has an equivalent form using only \(P_1\) and \(P_2\):

\[
P_1^T P_2 = 0
\]

\[
H_0 :
\]
Pattern Change Detection Based on Hypothesis Test (cont.)

- Maximum likelihood estimation with hypothesis $H_0$: $P_1^T P_2 = 0$

$$\mathcal{L}(P_2, S_2) = \|X_2 - P_2 S_2^T\|^2 + \lambda \|P_2^T P_1\|^2$$

- Maximum likelihood estimation without hypothesis $H_0$:

$$\mathcal{L}(P_2, S_2) = \|X_2 - P_2 S_2^T\|^2$$

- Likelihood ratio test:

$$\Lambda = \frac{\|X_2 - \hat{P} \hat{S}^T\|^2}{\|X_2 - \hat{P}_H \hat{S}_H^T\|^2}$$

Reject $H_0$ when $\Lambda < h$
Experiment summary

- **Text data sets**
  New combination of words – New topic

- **Face image data sets**
  New combination of pixels – New object

- **Surveillance videos**
  New combination of motion vectors – New event
Experimental results: synthetic change

<table>
<thead>
<tr>
<th>Name</th>
<th>Part 1</th>
<th>Part 2</th>
<th>Sample no.</th>
<th>Dim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>sys</td>
<td>comp.sys.ibm.pc</td>
<td>comp.sys.mac</td>
<td>400 × 2</td>
<td>1558</td>
</tr>
<tr>
<td>ossys</td>
<td>comp.os.ms-windows.misc,</td>
<td>comp.sys.mac, comp.windows.x</td>
<td>200 × 4</td>
<td>2261</td>
</tr>
<tr>
<td></td>
<td>comp.sys.ibm.pc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>computer</td>
<td>comp.graphics, comp.os.ms-windows.misc, comp.sys.ibm.pc</td>
<td>comp.sys.ibm.pc, comp.sys.mac, comp.windows.x, sci.electronics</td>
<td>100 × 6</td>
<td>1606</td>
</tr>
<tr>
<td>socialtalk</td>
<td>talk.politics.guns, talk.politics.mide</td>
<td>alt.atheism, assoc.religion.christian, talk.politics.mide, talk.religion.misc,</td>
<td>100 × 6</td>
<td>3312</td>
</tr>
<tr>
<td>sci</td>
<td>sci.crypt</td>
<td>sci.med</td>
<td>400 × 2</td>
<td>2870</td>
</tr>
<tr>
<td>rec-sci</td>
<td>rec.sport.baseball, rec.sport.hockey</td>
<td>sci.electronics, sci.space</td>
<td>200 × 4</td>
<td>2800</td>
</tr>
<tr>
<td>comp-sci</td>
<td>comp.graphics, comp.os.ms-windows.misc, comp.sys.ibm.pc</td>
<td>sci.electronics, sci.med, sci.space</td>
<td>100 × 6</td>
<td>1864</td>
</tr>
<tr>
<td>rec-talk</td>
<td>rec.autos, rec.motorcycles, rec.sport.baseball, rec.sport.hockey</td>
<td>talk.politics.guns, talk.politics.mide, talk.politics.mide, talk.religion.misc,</td>
<td>100 × 6</td>
<td>2992</td>
</tr>
</tbody>
</table>

The configuration of the pattern change data sets by using 20-news-group data sets.
Experimental results: synthetic change (cont.)

For each pair of red and blue bars, a smaller overlap in between indicates a better performance.
Experimental results: event detection from video

http://cs.binghamton.edu/~mrldata/Report
The Summary

- **Principal angles** to measure pattern change between high-dimensional data sets.

- **Matrix factorization** to summarize pattern space of high likelihood

- **Likelihood ratio test** based on linear model to unify the two tasks

- Experiments on text, images, and videos for justification.
Thank You!