Unsupervised Learning

Introduction

René Vidal
Center for Imaging Science
Johns Hopkins University
High-Dimensional Data

- In many areas, we deal with high-dimensional data
  - Computer vision
  - Medical imaging
  - Medical robotics
  - Signal processing
  - Bioinformatics
High-Dimensional Data in Computer Vision

High-Dimensional Data in Computer Vision

- 140 billion images
- 350 million new photos/day
- 3.8 trillion of photographs
- 10% in the past 12 months

http://www.buzzfeed.com/hunterschwarz/how-many-photos-have-been-taken-ever-6zgv

- 120 million videos
- 100 hours of video/minute
- 90% of the internet traffic will be video by the end of 2017
High-Dimensional Data in Computer Vision

- **ImageNet**: 14M images (1M with bounding box annotations), 22K categories

http://image-net.org
Big Data in Biomedical Imaging

• 400 million procedures/year involve at least 1 medical image

• Medical image archives are increasing by 20-40 percent each year

• 1 billion medical images stored in the US

• 1/3 of global storage is medical image information

• One individual’s online medical record could equate to 12 billion novels

Big Data in Biomedical Imaging

- High throughput neuroinformatics: bits of neuroscience at 1mm scale
  - 3000 brains
  - 1000x1000x500x100 dimensions
  - 1000-2000 relevant variables
Big Data in Biomedical Imaging

The Language of Surgery

Modeling the skills of human expert surgeons to train a new generation of students.  


L. Zapella, B. Bejar, R. Vidal. Surgical Gesture Classification from Video Data, MICCAI 2012 (Best paper Award).
How Do We Make Sense of Big Data?

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. XIII, 001–14 (1960)

The Unreasonable Effectiveness of Mathematics in the Natural Sciences

Richard Courant Lecture in Mathematical Sciences delivered at New York University, May 11, 1959

EUGENE P. WIGNER
Princeton University

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, Google
What is This Class About?

- Unsupervised learning methods for discovering structure in big, corrupted, high-dimensional data.
Course Information: Syllabus

• Introduction (Chapter 1)
• Part I: Single Subspace
  – Principal Component Analysis (Chapter 2)
  – Robust Principal Component Analysis (Chapter 3)
  – Kernel PCA and Manifold Learning (Chapter 4)
• Part II: Multiple Subspaces
  – Algebraic Methods (Chapter 5)
  – Statistical Methods (Chapter 6)
  – Spectral Methods (Chapter 7)
  – Sparse and Low-Rank Methods (Chapter 8)
• Part III: Applications
  – Image Representation (Chapter 8)
  – Image Segmentation (Chapter 9)
  – Motion Segmentation (Chapter 10)

http://link.springer.com.proxy1.library.jhu.edu
https://www.amazon.com/dp/0387878106/
Principal Component Analysis (PCA)

• Given a set of points lying in one subspace, identify
  – Geometric PCA: find a subspace S passing through them
  – Statistical PCA: find projection directions that maximize the variance

• Solution (Beltrami’1873, Jordan’1874, Hotelling’33, Eckart-Householder-Young’36)

\[ U \Sigma V^\top = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \end{bmatrix} \in \mathbb{R}^{D \times N} \]

• Applications:
  – Signal/image processing, computer vision (eigenfaces), machine learning, genomics, neuroscience (multi-channel neural recordings)
Application to Face Classification

**Problem:**
- Given face images with labels, use them to classify new face images

**Challenges:**
- Corruptions: occlusions, disguise
- Face detection
- Pose variations
- Light variations

**Subspace-based approaches:**
- Face images live in a subspace

Robust Principal Component Analysis

- Missing Entries

- Corrupted Entries

- Outliers
NonLinear PCA and Manifold Learning
Generalized Principal Component Analysis

- Given a set of points lying in multiple subspaces, identify
  - The number of subspaces and their dimensions
  - A basis for each subspace
  - The segmentation of the data points

- "Chicken-and-egg" problem
  - Given segmentation, estimate subspaces
  - Given subspaces, segment the data

- Challenges
  - Noise
  - Missing entries
  - Outliers
Applications of GPCA

- Face clustering and classification
- Motion segmentation
- DT segmentation
- Video segmentation

• Lossy image representation
Generalized Principal Component Analysis

- **Iterative methods**

- **Probabilistic methods**
  - Mixtures of PPCA (Tipping-Bishop ’99, Grubber-Weiss ’04, Kanatani ’04, Archambeau et al. ’08, Chen ’11)
  - Agglomerative Lossy Compression (Ma et al. ’07, Rao et al. ’08)
  - RANSAC (Leonardis et al.’02, Yang et al. ’06, Haralik-Harpaz ’07)

- **Algebraic methods**
  - Factorization (Boult-Brown’91, Costeira-Kanade’98, Gear’98, Kanatani et al.’01, Wu et al.’01)
  - Generalized PCA: (Shizawa-Maze ’91, Vidal et al. ’03 ’04 ’05, Huang et al. ’05, Yang et al. ’05, Derksen ’07, Ma et al. ’08, Ozay et al. ’10)