Laplacian Eigenmaps for Dimensionality Reduction and Data Representation

Lisa Wainer

Supervised by

Dr. Massi Pontil
Dimensionality Reduction

- Central problem in Machine Learning
- Arises in other fields
  - Information Processing
  - Data Compression
  - Scientific Visualisation
  - Pattern Recognition
Aims

- To obtain compact representation of data
  - Capturing desired information
  - Without loss of too much information
Benefits

- Data in a more compact form allows
  - Visualisation of the data
  - Speeds up processing time
  - Focus on significant features
Dimensionality Reduction as a Mapping
Methods in General Use

- Principle Components Analysis
  - Eigenvector method
  - Simple to implement

- PCA is limited to linear projections
  - Cannot produce good representation of non-linear data
A more powerful method

- Using Laplacian Eigenmaps
  - Belkin & Niyogi, 2001

- Unsupervised Learning algorithm

- Can generate non-linear mappings
  - Represent non-linear data
Algorithm Overview

- **Stage 1**
  - Builds a graph incorporating neighbourhood information of the data set

- **Stage 2**
  - Computes low dimensional representation
    - Optimally preserving local relationships between points

- Assumes the data sits on an underlying manifold
What is a Manifold?

- Non-linear lower dimensional substructure
  Eg. A 'Swiss Roll'
  - 2-D submanifold lying in 3-D space
Other Motivation

- Preserving local information seems to emphasise *natural clusters* in the data

- Possible role in human perception
  - Suggests a possible mechanism for emergent categorisation
Main Technical Aims

- Understand and Implement the algorithm
  - Matlab implementation

- Test the algorithm on data
  - Real (eg. images of faces)
  - Synthesised - see Toy example
  - Dependent on run time of Matlab implementation
Additional Technical Aims

- Time Permitting
- Study the use of the algorithm
  - Classification Problems & Clustering
    - Eg. Faces or Non-faces
- Comparisons with other non-linear methods
  - Kernal PCA
Plan

- 3 months to complete project

Outline
- Milestone 1: Algorithm implementation
- Milestone 2: Testing on Data
- Milestone 3: Writeup

Meetings with supervisor
- Twice a week when possible
- Begin June 3rd
Schedule

Understanding

Implementation I

Implementation II

Testing

Test Performance

Writeup
General Approach

- Iterative, incremental approach
  - Typical for scientific project

- Most resembles 'Agile' method (Collins-Cope)
  - Plan in detail for the short term
  - Broad strokes for long term
  - Risk analysis
Risk 1

- Computer resources
  - Effects possibility of testing on real data
    - May be able to negotiate with department
    - Focus on other additional Aims
Risk 2

- Supervisor absence
  - Aware that supervisor will be away
- Dates of holidays and plan accordingly
- Core understanding early on
- Possibly identify a second supervisor
Risk 3

- Running out of time
  - Disciplined approach to work
  - Keep regular hours

- Aim to do each week
  - Book work
  - Computer work
  - Written work
Risk Minimisation

- Prioritise
  - Understanding and Development of algorithm
  - Testing on Toy data set
  - Writing up as go along

- Maximise time with supervisor
  - More frequent meetings
  - Encourages work ethic
Metrics for Project Cycle

- Highly mathematical algorithm
  - Gives landmarks to assess progress

- Meeting milestones
  - Reports to supervisor
Metrics for Testing

- Test on Toy Example data set
- Compare with results from Belkin & Niyogi (2001)
  - Does my implementation perform similarly?
- Identified package available (Belkin)
Toy Example Data

Binary Images of Vertical and Horizontal Bars
Toy Example
Representations of the Data

From (Belkin & Niyogi, 2001)
Metrics for the Algorithm

- Can the overall method be assessed?
  - Comparison with other method
    - Kernal PCA : Non-linear form of PCA
      - Information loss
      - Runtime
Stakeholders

- Lisa Wainer
- Dr. Massi Pontil
- UCL CS Department
Resources

- PC & Matlab
- Toy Data Sets
  - Will generate
- Real data sets
  - Images of faces