COMP90051

Workshop Week 09

About the Workshops

- 7 sessions in total
  - Tue 12:00-13:00 AH211
  - Tue 12:00-13:00 AH108 *
  - Tue 13:00-14:00 AH210
  - Tue 16:15-17:15 AH109
  - Tue 17:15-18:15 AH236 *
  - Tue 18:15-19:15 AH236 *
  - Fri 14:15-15:15 AH211

About the Workshops

- Homepage
  - https://trevorcohn.github.io/comp90051-2017/workshops
- Solutions will be released on next Friday (a week later).
- No lecture & workshop next week

Syllabus

1	Introduction; Probability theory	Probabilistic models; Parameter fitting	
2	Linear regression; Intro to regularization	Logistic regression; Basis expansion	
3	Optimization; Regularization	Perceptron	
4	Backpropagation	CNNs; Auto-encoders	
5	Hard-margin SVMs	Soft-margin SVMs	
6	Kernel methods	Ensemble Learning	
7	Clustering	EM algorithm	
8	Principal component analysis; Multidimensional Scaling	Manifold Learning; Spectral clustering	←
9	Bayesian fundamentals	Bayesian inference with conjugate priors	
10	PGMs, fundamentals	Conditional independence	
11	PGMs, inference	Belief propagation	
12	Statistical inference; Apps	Subject review	

Outline

- Review the lecture, background knowledge, etc.
  - Principal component analysis (PCA)
  - Multidimensional scaling (MDS)
  - Isomap
    - Geodesic distance
  - Spectral clustering
    - Laplacian eigenmaps
- Work on project 2

Principal components analysis

- Input
  - N points (high dimensional)
- Output
  - Components with their variance
  - N points (low dimensional)
- Objective
  - Find maximum variance subspace

Multidimensional scaling (MDS)

- Input
  - Distance matrix D ($N \times N$)
- Output
  - N points in d dimensions
- Objective
  - Find N points in d dimensions
  - So that their pair-wise distances are similar to those in D

Euclidian distance + MDS

- Input
  - N points (high dimensional)
- Output
  - N points (low dimensional)
- Objective
  - Transform N points to lower dimension
  - So that their pair-wise Euclidian distances are preserved

Isomap = geodesic distance + MDS

- Input
  - N points (high dimensional)
- Output
  - N points (low dimensional)
- Objective
  - Transform N points to lower dimension
  - So that their pair-wise geodesic distances are preserved

Geodesic distance

- Construct a graph
  - There are edges between neighbouring points
    - Option 1: distance $<$ a threshold
    - Option 2: k-nearest neighbours
  - Weights are distances
- Find all-pairs shortest paths
- The geodesic distance is the length of the shortest path

Laplacian eigenmaps

- Input
  - N points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ (high dimensional)
- Output
  - N points $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N$ (low dimensional)
- Objective
  - Similar points are closer in the lower dimensional space

$$\min_{\boldsymbol{z}} \sum_{i,j} \left\| \boldsymbol{z}_i - \boldsymbol{z}_j \right\|^2 A_{ij}$$

  - where A is the adjacency matrix

Adjacency matrix

- Adjacency matrix A for a graph
  - Option 1&2

$$A_{ij} = \begin{cases} 1 & \text{if } \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ are neighbouring points} \\ 0 & otherwise \end{cases}$$

  - Option 3

$$A_{ij} = \exp\left(-\frac{1}{\sigma}\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2\right)$$

- A kind of similarity matrix

Spectral clustering

- Laplacian eigenmaps + clustering

- Laplacian eigenmaps for dimension reduction
- Any clustering method, for example, k-means

A summary

- Dimension reduction
  - PCA
  - Euclidian distance + MDS
  - Isomap = geodesic distance + MDS
  - Laplacian eigenmaps
- Clustering
  - Spectral clustering = Laplacian eigenmaps + clustering

Outline

- Review the lecture, background knowledge, etc.
  - Principal component analysis (PCA)
  - Multidimensional scaling (MDS)
  - Isomap
    - Geodesic distance
  - Spectral clustering
    - Laplacian eigenmaps

- Work on project 2