Homework 5 Using Nonlinear Mapping for Data Structure Analysis

(300 Points)

Assigned Date: Wednesday, November 18, 2009

		Due Date
Oral Presentation	100 points	4:00 PM Wednesday November 25
Phase I	100 points	4:00 PM Monday December 7
Phase II	100 points	4:00 PM Monday December 14

Educational Goal

To develop skills of using Artificial Intelligence algorithms for solving practical problems.

Description

In this assignment, we will understand, implement, and improve the nonlinear mapping algorithm proposed by the following paper:

"*A Nonlinear Mapping for Data Structure Analysis*," IEEE Transactions on Computers archive, Volume 18 , Issue 5 (May 1969), Pages 401-409, J. W. Sammon.

ABSTRACT

An algorithm for the analysis of multivariate data is presented along with some experimental results. The algorithm is based upon a point mapping of N L-dimensional vectors from the L-space to a lower-dimensional space such that the inherent data "structure" is approximately preserved.

Oral Presentation

4:00 – 5:15 PM November 25

Requirements

1. Read Sammon's paper and present your understanding of the paper in class.

Submission Requirements

1. Prepare Microsoft PowerPoint slides for your presentation.
2. Save the PowerPoint slides as AI_ lastname_firstname_hw5. For example, student John Smith should name his file as AI_Smith_John_hw5. Submit the softcopy through your UMassOnline account at http://boston.umassonline.net/index.cfm.
3. Prepare a paper copy of the PowerPoint slides with 2 slides per page. Submit the paper copy to the instructor on the presentation day.
4. **Presentation duration**. You will have 10 minutes on presentation. We will have 20-minute for Q&A session for all the presenters.
5. **Presentation content**. Because audiences understand the general idea of the paper, the presentation should focus on your own understanding and comments on the paper.

Phase I Implementing the NLM Algorithm

Requirements

- Implement the Nonlinear Mapping Algorithm (NLM) proposed in Sammon's paper. The pseudo code of the algorithm is given in Appendix I.
- Use Wisconsin Diagnostic Breast Cancer (WDBC) as the input data set.
- Generate Sammon's Map in 2 dimensions of the breast cancer data set. Plot the resulting Sammon's Map.
- Report the number of iterations and the resulting error value of the Sammon's Map.

Submission Requirements

1. Follow the language requirements for programming assignments posted at http://www.cs.umb.edu/~ding/classes/470_670/student.htm
2. Your program should be well-documented. Variable names and function names should be self-descriptive. Major functions should be explained clearly in comments.
3. Turn in the paper copy and soft copy of all the files **including your program code and results**. Submit **a single zipped file** of all the files of this assignment through your UMassOnline account at http://boston.umassonline.net/index.cfm. Submit the paper copy along with the cover page in class. Paper copy should be bound firmly together as one pack (for example, staple, but not limited to, at the left corner). 5 points will be deducted for unbounded homework.
4. Name your file with AI_ lastname_firstname_hw5_phaseOne. For example, student John Smith should name his file as AI_Smith_John_hw5_phaseOne.
5. No hard copies or soft copies results in 0 points.

Phase II Improving the NLM Algorithm

Requirements

- The algorithm used by Sammon is a type of hill-climbing local search algorithm. Improve the algorithm using the simulated annealing algorithm.
- Use Wisconsin Diagnostic Breast Cancer (WDBC) as the input data set; generate Sammon's Map in 2 dimensions of the breast cancer data set, plot the resulting Sammon's Map; report the number of iterations and the resulting error value of the Sammon's Map.
- Compare the result of Phase II and Phase I and conclude which algorithm can produce better result.

Submission Requirements

1. Follow the language requirements for programming assignments posted at http://www.cs.umb.edu/~ding/classes/470_670/student.htm
2. Your program should be well-documented. Variable names and function names should be self-descriptive. Major functions should be explained clearly in comments.
3. Turn in the paper copy and soft copy of all the files **including your program code and results**. Submit **a single zipped file** of all the files of this assignment through your UMassOnline account at http://boston.umassonline.net/index.cfm. Submit the paper copy along with the cover page in class. Paper copy should be bound firmly together as one pack (for example, staple, but not limited to, at the left corner). 5 points will be deducted for unbounded homework.
4. Name your file with AI_ lastname_firstname_hw5_phaseTwo. For example, student John Smith should name his file as AI_Smith_John_hw5_phaseTwo.
5. No hard copies or soft copies results in 0 points.