Two hours - online hybrid

EXAM PAPER MUST NOT BE REMOVED FROM THE EXAM ROOM

UNIVERSITY OF MANCHESTER
DEPARTMENT OF COMPUTER SCIENCE

Modelling and visualisation of high-dimensional data

Date: Tuesday 21st January 2020

Time: 09:45 - 11:45

This is a hybrid examination with sections to be answered online and questions to be answered on paper
The examination is worth a total of 50 marks

Please answer All Questions in Section A online
and All Questions in Section B in a separate answerbook

This is a CLOSED book examination

Electronic calculators may be used in accordance with the University regulations

Section A

Section A contains restricted multiple choice questions (MCQs) and is NOT published

Section B

Answer *all* questions from this section.

1. *Multi-dimensional scaling* (MDS) refers to a class of learning algorithms that can provide a visual representation of the pattern of proximities among a set of objects. Describe the following terminologies and techniques used in MDS:

   a) What is the *disparity*? (2 marks)

   b) What is the *stress*? What is the role that a stress plays in MDS? (2 marks)

   c) What is the stress used in one of the most popular MDS algorithms, *Sammon nonlinear mapping*? (2 marks)

2. The *Locally Linear Embedding* (LLE) is a typical non-linear manifold learning algorithm.

   a) Give all the cost functions used to derive the LLE algorithm with the notation used in lecture notes and explain their roles in terms of manifold learning. (4 marks)

   b) State the hyper-parameter in the LLE learning algorithm and explain the adversarial effect incurred by choosing an improper value. (2 marks)

3. For a data set of N data points in a high-dimensional space, $X = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$, we would like to find a *single* data point, a zero-dimensional representation of this data set, such that the sum of the squared distances between this point and all the data points in the data set is as small as possible. Prove such a data point must be the mean of N data points in this data set. (8 marks)

END OF EXAMINATION