COMP/ENGN8535 Homework Assignment #3

General Instructions

- Due Date: Sunday 23 April 2023, 23:59pm (AEST).
- Full marks for this homework = 100 points.
- Please complete as many problems as you can, and answer the questions clearly and concisely. In all your solutions, you need to show the key steps for how you reach your answers. The clarity of your answers may affect the final marks.
- Submit your answers in a single "pdf" file, with file name U4056789-HW-1.pdf (Note: replace the U4056789 with your own university ID) to Wattle before the deadline. It is a good idea to also include a cover page with your full name and university ID so that we can identify your submission easily when marking.
- In preparing your PDF submission, both typed and handwritten-and-scanned answers are acceptable, though the former is preferred due to its clarity.
- Late submissions will not be accepted and incur a Mark of 0.

Academic Integrity: Your homework must be completed individually and independently. You must comply with the University Policy on Academic Integrity. Plagiarism will lead to severe consequences, including academic misconduct. You are allowed to discuss the problems and solution ideas with other students, but your final solutions must be your own work and written in your own language.

Problems

1. (20 points) Consider the following dataset of 4 data points:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \; \mathbf{x}_3 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \; \mathbf{x}_4 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

   (a) (10 points) Compute the principle components of the data using PCA.

   (b) (10 points) Compute scalar output data points $\{y_j \in \mathbb{R} : 1 \leq j \leq 4\}$ using Kernel PCA with a polynomial kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)^\gamma$ where $\gamma = 10$.

2. (50 points) Consider a data matrix $X \in \mathbb{R}^{d \times n}$ whose columns are d-dimensional data points that have already had their mean subtracted (the mean vector of the data points in X is $\mu = 0$).

   (a) (20 points) The matrix $C = \dfrac{1}{n} X X^T \in \mathbb{R}^{d \times d}$ is the data covariance matrix and $\kappa = \dfrac{1}{n} X^T X \in \mathbb{R}^{n \times n}$ is the data Gram matrix. Show that

$$\max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u = \max_{\{v \in \mathbb{R}^n : \|v\|_2 = 1\}} v^T \kappa v.$$

   (b) (15 points) With reference to your answer to Part (a), explain how PCA can be performed using the Gram matrix κ instead of the covariance matrix C.

   (c) (15 points) With reference to your answer to Part (b), show that Kernel PCA reduces to (standard) PCA with the choice of the kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$.

3. (20 points) Consider the following dataset of 4 data points:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \; \mathbf{x}_3 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \; \mathbf{x}_4 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

   (a) (5 points) Compute the matrix $D \in \mathbb{R}^{4 \times 4}$ of (squared) distances between the data points (i.e., the components of D should satisfy $D_{ij} = \|\mathbf{x}_j - \mathbf{x}_i\|^2$).

   (b) (15 points) Perform Multidimensional Scaling (MDS) on the matrix D you computed in Part (a) to find scalar output data points $\{y_j \in \mathbb{R} : 1 \leq j \leq 4\}$.

4. (10 points) IsoMap and Locally Linear Embedding (LLE) are two popular approaches to manifold learning. What advantages does LLE have compared to IsoMap in terms of computational and memory requirements?

END.