MLAI Week 8 Exercise: Unsupervised Learning

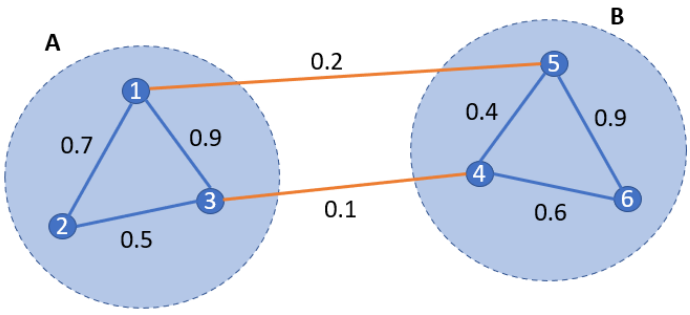
Note: An indicative mark is in front of each question. The total mark is 13. You may mark your own work when we release the solutions.

[1] 1. Consider 30-bit deep colour images of size 1200×1200. How many possible images of this size and bit depth are there?

[2] 2. We are using PCA to reduce data dimensionality from 3 to 2. The top two eigenvectors are $\begin{pmatrix} 0.4729 & -0.8817 \\ -0.8817 & -0.4719 \\ 0 & 0 \end{pmatrix}$ where each column is an eigenvector. Use this PCA transformation to reduce the dimensionality of two data points $\mathbf{x}_1 = (2, 3, 3)^\top$ and $\mathbf{x}_2 = (4, 1, 0))^\top$ to 2 as $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$. Show the procedures to compute $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$.

[2] 3. Given a dataset $\{0, 2, 4, 6, 24, 26\}$, initialise the k-means clustering algorithm with 2 cluster centres $c_1 = 3$ and $c_2 = 4$. What are the values of c_1 and c_2 after one iteration of k-means? What are the values of c_1 and c_2 after the second iteration of k-means?

[2] 4. For the graph below, compute the normalised cut $Ncut(A, B)$.



[3] 5. An alternative to derive PCA is to minimize the reconstruction error (Slide 26) for all N data samples $\mathbf{x}^{(i)}, i = 1, \cdots, N$, assuming that the mean $\boldsymbol{\mu} = \sum_i \mathbf{x}^{(i)}$ is zero. Take this approach to derive the first principal component (as the first eigenvector of the data matrix).

[3] 6. In spectral clustering, show that the smallest eigenvalue for the formulated generalized eigenvalue problem on Slide 41 is 0 with the corresponding generalized eigenvector $\mathbf{y} = \mathbf{1}$, hence the same "representation/embedding" for all nodes.