Week 9 Exercise Sheet

The following exercises have different levels of difficulty indicated by (*), (**), (***). An exercise with (*) is a simple exercise requiring less time or effort to solve compared to an exercise with (***), which is a more complex exercise.

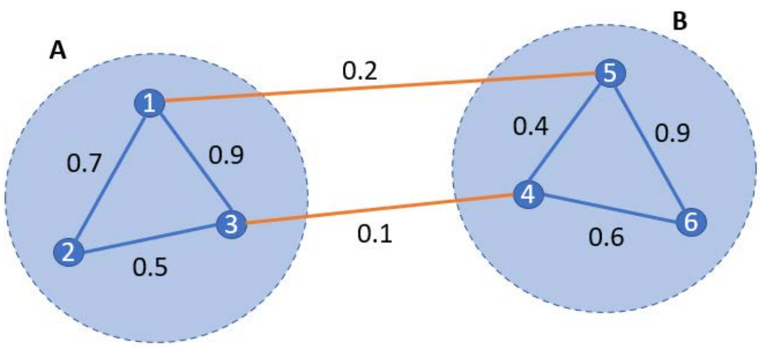
Unsupervised Learning

1. (**) Consider the following dataset:

$$\mathbf{x}_1 = (1, 1) \qquad \mathbf{x}_2 = (2, 2)$$
$$\mathbf{x}_3 = (3, 1) \qquad \mathbf{x}_4 = (4, 2)$$
$$\mathbf{x}_5 = (5, 1) \qquad \mathbf{x}_6 = (6, 2)$$

   Perform the K-means algorithm on this data to find 2 clusters. Initialise your centroids to $\mathbf{m}_1 = (0, 0)$ and $\mathbf{m}_2 = (7, 2)$, which datapoints are assigned to each cluster in the first iteration? What are the values of the centroids after the first iteration and then after the second iteration?

2. (**) For the graph below, compute the normalised cut, $\mathrm{Ncut}(A, B)$.



3. (***) In spectral clustering, the graph partitioning is solved through a generalised eigenvalue equation of the graph Laplacian

$$(\mathbf{D} - \mathbf{W})\mathbf{y} = \lambda \mathbf{D}\mathbf{y} \tag{1}$$

where $\mathbf{W}$ is the graph connection matrix, $\mathbf{D}$ is the degree matrix with diagonal entries $D_{ii} = d_i = \sum_j W_{ij}$. Show that $\mathbf{y} = \mathbf{1}$ (a vector of all ones) is an eigenvector of this equation and that its eigenvalue is $\lambda = 0$. What is the significance of this solution?