MS1b: SDM - Problem Sheet 1

1. (Biplot) Examine and use the R code from lecture 1 to create a biplot of the *state* data. In this question, you will interpret it.

   (a) According to the plot, what variables are positively correlated with graduating high school *HS Grad*? Which are negatively correlated? In each case, give a possible explanation.

   (b) Within the range of the data, you should find areas of voids, where relatively few states appear. Where are they and what does each one mean?

   (c) The data distribution also has a few "funnels", directions where the states become increasingly compact. What are they and what does each one mean?

   (d) If the eigenvalues associated with the first two eigenvalues explained a smaller amount of the total variance, what would be the danger in interpreting the plot as you did above?

2. (MDS) Under the assumption that your data are centred, show that you can compute the $n \times n$ Gram matrix B such that $b_{ij} = x_i^T x_j$ using the dissimilarity matrix D where $d_{ij} = \|x_i - x_j\|_2$.

3. (MDS) Give a sufficient condition on an $n \times n$ dissimilarity matrix D so that there is a representation of D in terms of a set of n points in two dimensions with the property that the Euclidean distance between points x_i and x_j exactly equals D_{ij}.

4. (SVD) Determine SVDs of the following matrices (by hand calculation):

$$\text{(a)} \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} \text{(b)} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \text{(c)} \begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \text{(d)} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \text{(e)} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

5. (PCA) Show that $\text{tr}(\tilde{V}_q^T X^T X \tilde{V}_q)$ is maximized (up to permutation) over matrices $\tilde{V}_q = (\tilde{v}_1, \ldots, \tilde{v}_q) \in \mathbb{R}^{p \times q}$ with q orthonormal columns (i.e. $\tilde{V}_q^T \tilde{V}_q = 1$) by $\tilde{v}_i = v_i$, where v_i, $i = 1, \ldots, q$ are the q eigenvectors of $X^T X v_i = \lambda_i v_i,\ i = 1, 2...q$ corresponding to the first (i.e. the largest) q eigenvalues of $X^T X$.