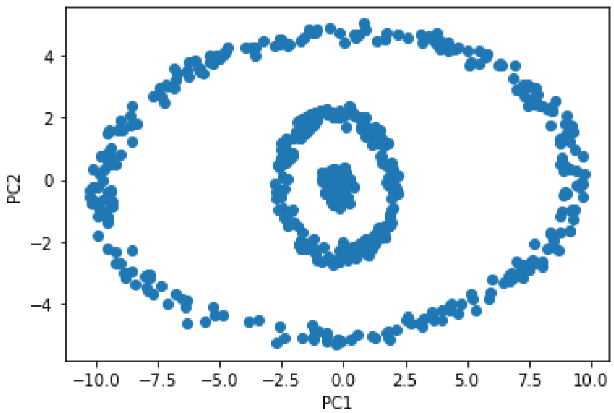
PHYS598 SDA Homework 2

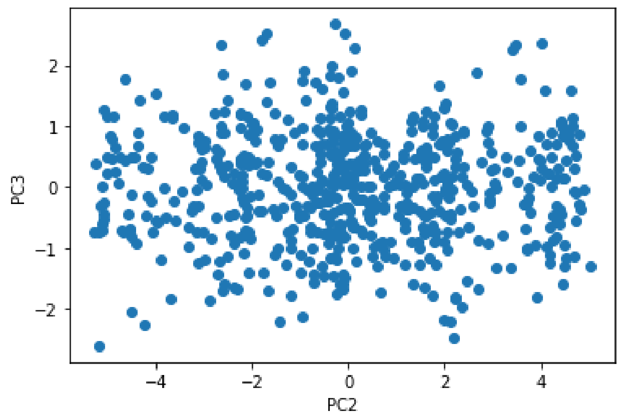
February 16, 2024

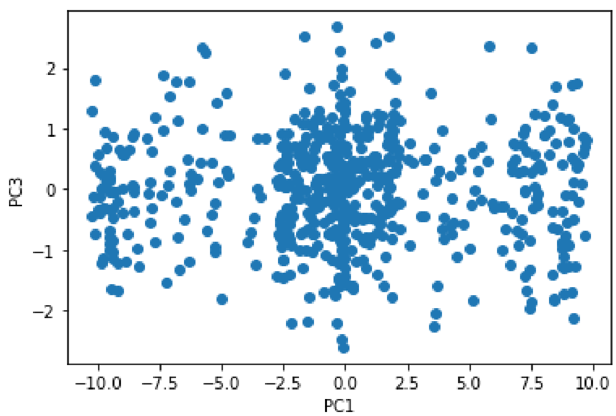
Problem 1

(a) Please see example code for a possible solution to this problem.

(b) The result of projecting onto the first three principal component directions yield the following plots.







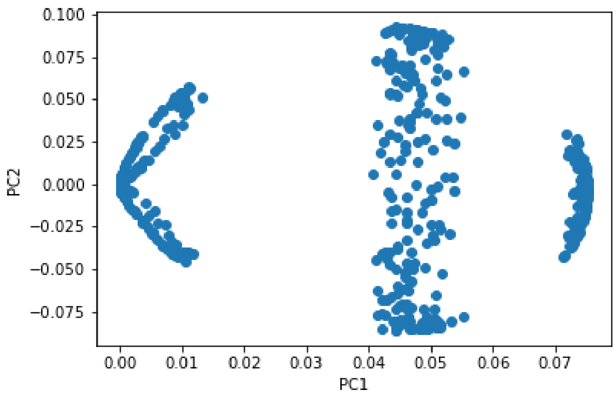
(c) The explained variance can be calculated using the output eigenvalues from PCA. In particular, the eigenvalues of the empirical covariance matrix Σ are σ_i^2, the variance associated to the i-th principcal component. The variance explained is then

$$\sigma^2_{\text{explained}} = \frac{\sum_{i=1}^{3} \sigma_i^2}{\sum_{i=1}^{600} \sigma_i^2} = \frac{\sum_{i=1}^{3} \sigma_i^2}{\text{tr}(\Sigma)} = \frac{\sum_{i=1}^{3} \sigma_i^2}{\|M\|_F^2} \tag{1}$$

where M is the data matrix. For the given data set, the result is

$$\sigma^2_{\text{explained}} \approx 0.1696 \tag{2}$$

(d) Once again, see the accompanying python file for an example solution. The results of kernel PCA with the radial basis function are displayed in the following plot.



These are now linearly clustered; however, K-means still fails to pick out three clusters reliably, as is shown below. This is because the three clusters are not spherical, nor are they of similar size.

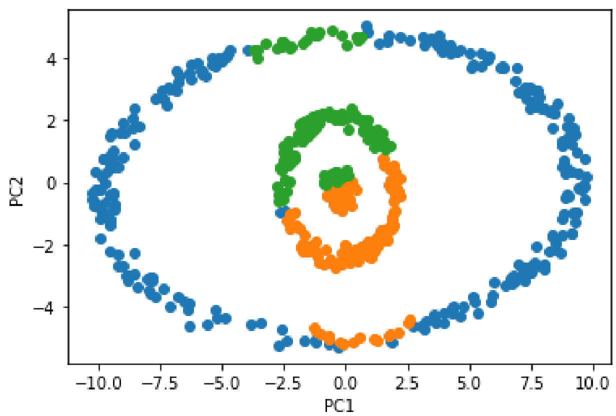


Figure 2: Result of K-means clustering with $k = 3$, with cluster labels assigned to the original PC1 vs. PC2 plot. Results may differ depending on the initialization of your implementation.

Problem 2

With the distance function $\delta_\alpha(s_i, s_j) = \delta(s_i, s_j) + \alpha(1 - \delta_{ij})$, the new distance squared matrix H' is given by

$$H'_{ij} = \delta^2_\alpha(s_i, s_j) = \delta^2(s_i, s_j) + 2\alpha\delta(s_i, s_j) + \alpha^2(1 - \delta_{ij}) \tag{3}$$

where we have used the fact that $\delta(s_i, s_j)(1 - \delta_{ij}) = \delta(s_i, s_j)$ and $(1 - \delta_{ij})^2 = (1 - \delta_{ij})$. Now define $D_{ij} = \delta(s_i, s_j)$ and $\Pi_{ij} = 1 - \delta_{ij}$. Then we have

$$H'_{ij} = H_{ij} + 2\alpha D_{ij} + \alpha^2 P_{ij} \tag{4}$$

But notice that $B = -\frac{1}{2}JHJ$ is linear in H, and so we have

$$B'_{ij} = B_{ij} + 2\alpha B^D_{ij} + \alpha^2 B^\Pi_{ij} \tag{5}$$

where $B = -\frac{1}{2}JHJ$, $B^D = -\frac{1}{2}JDJ$ and $B^\Pi = -\frac{1}{2}J\Pi J$. It is worthwhile to construct B^Π explicitly, which yields

$$B^\Pi_{ij} = -\frac{1}{2}\left(\Pi_{ij} - \Pi_{i\cdot} - \Pi_{\cdot j} + \Pi_{\cdot\cdot}\right) \tag{6}$$

$$= -\frac{1}{2}\left(\Pi_{ij} - 2\frac{m-1}{m} + \frac{m(m-1)}{m^2}\right) \tag{7}$$

$$= \delta_{ij} - \frac{1}{m} \tag{8}$$

This matrix is easily diagonalized. It has one zero eigenvector, namely $\vec{1} = (1, 1, \cdots, 1)$. Then, any other vector v such that $\vec{1}^t v = 0$ is also an eigenvector with eigenvalue 1. We see then that B^Π_{ij} is positive semi-definite, and is simply the identity operator when projected onto the subspace orthogonal to $\vec{1}$. Now we wish to consider under what conditions B' is positive semidefinite. First note that $J\vec{1} = \vec{1}^t J = 0$, and so it is sufficient to consider vectors v whose projection onto $\vec{1}$ is zero. For any such v, we have

$$v^t B' v = v^t B v + 2\alpha v^t B^D v + \alpha^2 \|v\|_2^2 \tag{9}$$

where we have used the fact that B^Π is the identity when projected onto the subspace orthogonal to $\vec{1}$. So the condition that B' is positive semi-definite reduces to

$$\hat{v}^t B \hat{v} + 2\alpha \hat{v}^t B^D \hat{v} + \alpha^2 \geq 0 \tag{10}$$

where we have divided through by $\|v\|_2^2$ to replace v with the normalized vector $\hat{v} = v/\|v\|_2$. So our search over all v has now reduced to a search over unit vectors $\hat{v}$ such that $\hat{v} \cdot \vec{1} = 0$. We may now use the fact that B and B^D are bounded linear operators (since we are working in finite dimensions) to define the following quantities:

$$L_1 = \min_{\substack{\|v\|_2=1 \\ v\cdot\vec{1}=0}} v^t B v \tag{11}$$

$$L_2 = \min_{\substack{\|v\|_2=1 \\ v\cdot\vec{1}=0}} v^t B^D v \tag{12}$$

$$\tag{13}$$

Then (10) will be satisfied for all desired v if

$$L_1 + 2\alpha L_2 + \alpha^2 \geq 0 \tag{14}$$

Using these values simultaneously represent the worst case scenario when L_1 and L_2 are simultaneously minimized by a single v. This is a simple system to solve, and in particular, we find that B' will be positive semi-definite so long as α is larger than the larger root of the quadratic:

$$\alpha \geq -L_2 + \frac{1}{2}\sqrt{L_2^2 - L_1} \tag{15}$$

For the sake of completeness, note that the RHS of the above inequality is always a real number if $L_1 \leq 0$. However if $L_1 > 0$, then B would have been positive semi-definite from the start. Thus, for $\alpha \geq \alpha_0 = -L_2 + \frac{1}{2}\sqrt{L_2^2 - L_1}$ and a matrix B_{ij} which is not positive semi-definite, the matrix B'_{ij} will be positive semi-definite, as desired.

Problem 3

(a) To determine if K is positive semi-definite we need to consider $v^t K v$ for some vector $v \in \mathbb{R}^m$ with $v \neq 0$. Written out, we have

$$\sum_{i,j=1}^{m} v_i K_{ij} v_j = \sum_{i,j}^{m} v_i \langle x_i, x_j \rangle v_j \tag{16}$$

Using linearity of the inner product, we may define the vector $w = \sum_i v_i x_i \in V$. Then we have

$$v^t K v = \langle w, w \rangle \geq 0 \tag{17}$$

Thus, K is positive semi-definite.

(b) K will be positive definite if and only if all the x_i are linearly independent. First let us consider the forward implication, which we prove by contradiction. If K is positive definite, then for any $v \neq 0$ we have

$$v^t K v > 0 \tag{18}$$

Now suppose we had some particular pair of x_i and x_j which were linearly dependent, where we may then write $x_j = \lambda x_i$. Then for $v_a = \delta_{ai} - \frac{1}{\lambda}\delta_{aj}$ we have

$$v^t K v = \langle x_i - \frac{1}{\lambda} x_j, x_i - \frac{1}{\lambda} x_j \rangle = \langle 0, 0 \rangle = 0 \tag{19}$$

and so we have a contradiction. Thus, all the x_i must be linearly independent.

For the converse statement the proof is straightforward. If all the x_i are linearly independent then they form a basis on some (possible proper) subspace of the inner product space. The linear combination $w = \sum_i v_i x_i$ then uniquely defines a non-zero vector so long as $v_i \neq 0$ for at least one i. But since $\langle w, w \rangle > 0$ if $w \neq 0$, we have shown that K must be positive definite.

Problem 4

As was shown in lecture, the solution to the approximate MDS embedding into $\mathbb{R}^n$ of the $m = 4$ data points is given by the $m \times n$ matrix X where

$$X_{:i} = v_i \sqrt{\max(0, \lambda_i)} \tag{20}$$

where λ_i and v_i are the eigenvalues and eigenvectors of B ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. For this particular problem, B is given by

$$B = \frac{1}{16} \begin{pmatrix} 15 & 1 & -17 & 1 \\ 1 & 3 & 1 & -5 \\ -17 & 1 & 15 & 1 \\ 1 & -5 & 1 & 3 \end{pmatrix} \tag{21}$$

The eigensystem of this matrix is given by

$$\lambda_1 = 2, \qquad \lambda_2 = \frac{1}{2}, \qquad \lambda_3 = 0, \qquad \lambda_4 = -\frac{1}{4} \tag{22}$$

$$v_1 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{pmatrix}, \qquad v_4 = \begin{pmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{pmatrix} \tag{23}$$

(a) For the embedding into $\mathbb{R}^1$, we simply need the maximal eigenvector, and so

$$X = \begin{pmatrix} -\sqrt{2} \\ 0 \\ \sqrt{2} \\ 0 \end{pmatrix} \tag{24}$$

The data points are mapped to the rows of X, and so $s_1 \to -\sqrt{2}$, $s_2 \to 0$, $s_3 \to \sqrt{2}$ and $s_4 \to 0$.

(b) For embedding in $\mathbb{R}^2$, we have

$$X = \begin{pmatrix} -\sqrt{2} & 0 \\ 0 & -\frac{1}{2\sqrt{2}} \\ \sqrt{2} & 0 \\ 0 & \frac{1}{2\sqrt{2}} \end{pmatrix} \tag{25}$$

And so $s_1 \to (-\sqrt{2}, 0)$, $s_2 \to (0, -\frac{1}{2\sqrt{2}})$, $s_3 \to (\sqrt{2}, 0)$ and $s_4 \to (0, \frac{1}{2\sqrt{2}})$.