SF2704 Clustering and persistence
Homework 3

Euclidean distance and scalar product in R^k.

Recall:

- Let $x = (x_1, \dots, x_k)$ and $y = (y_1, \dots, y_k)$ be points in $\mathbf{R}^k$. The symbol $\vec{xy}$ denotes the vector in $\mathbf{R}^k$ whose coordinates are given by:

$$\vec{xy} = \begin{bmatrix} y_1 - x_1 \\ y_2 - x_2 \\ \vdots \\ y_k - x_k \end{bmatrix}$$

- The Euclidean scalar product of two vectors $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}$ in $\mathbf{R}^k$ is given by:

$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 + y_2 + \cdots + x_k y_k$$

- An Euclidean distance between two points $x = (x_1, \dots, x_k)$ and $y = (y_1, \dots, y_k)$ in $\mathbf{R}^k$ is given by:

$$d_e(x, y) = \sqrt{\vec{xy} \cdot \vec{xy}} = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots (y_k - x_k)^2}$$

  This is a metric on $\mathbf{R}^k$.

Exercise 0.

Show that for any three points x_0, y, and z in $\mathbf{R}^k$:

$$\vec{x_0 y} \cdot \vec{x_0 z} = \frac{1}{2}(d_e(y, x_0)^2 + d_e(z, x_0)^2 - d_e(y, z)^2)$$

Conclude that from the Euclidean distance between points we can recover the Euclidean scalar product of vectors.

Positive and non-negative definite symmetric matrices.

Recall:

- For any symmetric $n \times n$ matrix A, there is an orthogonal $n \times n$ matrix B (orthogonal means that $BB^t = I$), such that:

$$B^{-1}AB = B^tAB = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

  where λ_i are eigenvalues of A.

- A symmetric $n \times n$ matrix A is called **positive definite** if all its eigenvalues are strictly bigger than 0.

- A symmetric $n \times n$ matrix A is called **non-negative definite** if all its eigenvalues are bigger or equal than 0.

Exercise 1.

Let $\vec{v}_1, \ldots, \vec{v}_n$ be vectors in $\mathbf{R}^k$. Show that the following $n \times n$ matrix is non-negative definite:

$$\begin{bmatrix} v_1 \cdot v_1 & v_1 \cdot v_2 & \cdots & v_1 \cdot v_n \\ v_2 \cdot v_1 & v_2 \cdot v_2 & \cdots & v_2 \cdot v_n \\ \vdots & \vdots & & \vdots \\ v_n \cdot v_1 & v_n \cdot v_2 & \cdots & v_n \cdot v_n \end{bmatrix} = \begin{bmatrix} v_1^t \\ v_2^t \\ \vdots \\ v_n^t \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

Prove that this matrix is positive definite if and only if the vectors $\vec{v}_1, \ldots, \vec{v}_n$ are linearly independent.

Exercise 2.

Let A be a symmetric $k \times k$ matrix. Define $(\vec{x}, \vec{y})_A = \vec{x}^t A \vec{y}$. Prove that the following are equivalent:

- there is a linear isomorphism $f\colon \mathbf{R}^k \to \mathbf{R}^k$ such that $\vec{x} \cdot \vec{y} = (\vec{x}, \vec{y})_A$ for any vectors $\vec{x}$ and $\vec{y}$ in $\mathbf{R}^k$.

- A is positive definite.

(Hint: use the matrix B and $\begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_k} \end{bmatrix}$ to construct f.)

Conclude that if A is positive definite then, for any vector $\vec{x} \neq 0$ in $\mathbf{R}^k$, $(\vec{x}, \vec{x})_A > 0$.

Exercise 3.

Let A be a positive definite symmetric $k \times k$ matrix. Let x and y be points in $\mathbf{R}^k$. Define $d_A(x, y) = \sqrt{(\overrightarrow{xy}, \overrightarrow{xy})_A}$. Prove that d_A is a metric on $\mathbf{R}^k$. Use exercise 2 to show that $\mathbf{R}^k$ with the Euclidean metric is isometric to $\mathbf{R}^k$ with the metric d_A.

Embeddings into Euclidean spaces.

Let $X = \{x_0, \ldots x_n\}$ be a finite set and d be a metric on X.

- We say that X embeds into an Euclidean space if there is a function $f\colon X \to \mathbf{R}^k$ such that $d(x, y) = d_e(f(x), f(y)$ for any x and y in X.
- We can think about x_0 as the origin and about pairs of points $x_0 y$ as vectors and denote them by $\overrightarrow{x_0 y}$. Inspired by exercise 0 we can use the metric d to define a scalar product on vectors as:

  $$(\overrightarrow{x_0 y}, \overrightarrow{x_0 z}) = \frac{1}{2}(d(y, x_0)^2 + d(z, x_0)^2 - d(y, z)^2)$$

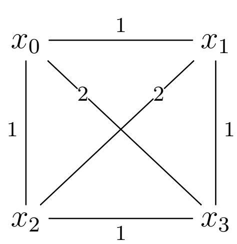
  This can be used this to define an $n \times n$ matrix:

  $$D := \begin{bmatrix} (\overrightarrow{x_0 x_1}, \overrightarrow{x_0 x_1}) & (\overrightarrow{x_0 x_1}, \overrightarrow{x_0 x_2}) & (\overrightarrow{x_0 x_1}, \overrightarrow{x_0 x_3}) & \cdots & (\overrightarrow{x_0 x_1}, \overrightarrow{x_0 x_n}) \\ (\overrightarrow{x_0 x_2}, \overrightarrow{x_0 x_1}) & (\overrightarrow{x_0 x_2}, \overrightarrow{x_0 x_2}) & (\overrightarrow{x_0 x_2}, \overrightarrow{x_0 x_3}) & \cdots & (\overrightarrow{x_0 x_2}, \overrightarrow{x_0 x_n}) \\ (\overrightarrow{x_0 x_3}, \overrightarrow{x_0 x_1}) & (\overrightarrow{x_0 x_3}, \overrightarrow{x_0 x_2}) & (\overrightarrow{x_0 x_3}, \overrightarrow{x_0 x_3}) & \cdots & (\overrightarrow{x_0 x_3}, \overrightarrow{x_0 x_n}) \\ \vdots & \vdots & \vdots & & \vdots \\ (\overrightarrow{x_0 x_n}, \overrightarrow{x_0 x_1}) & (\overrightarrow{x_0 x_n}, \overrightarrow{x_0 x_2}) & (\overrightarrow{x_0 x_n}, \overrightarrow{x_0 x_3}) & \cdots & (\overrightarrow{x_0 x_n}, \overrightarrow{x_0 x_n}) \end{bmatrix}$$

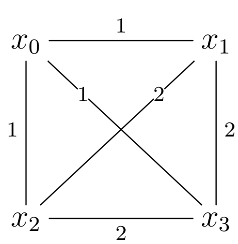
Exercise 4.

1. Prove that the matrix D above is symmetric.
2. Calculate this matrix for the following metrics and decide if it is positive definite or non-negative definite:

(a)



(b)



Exercise 5.

Show that if (X, d) embeds into en euclidean space, then the matrix D is non-negative definite.

Exercise 6.

Assume that the matrix D is positive definite. Define $f\colon X \to \mathbf{R}^n$ by the formula:

$$f(y) = ((y\vec{x}_0, y\vec{x}_1), (y\vec{x}_0, y\vec{x}_2), \ldots, (y\vec{x}_0, y\vec{x}_n))$$

Show that $d(x, y) = d_D(f(x), f(y))$. Conclude that (X, d) can be embedded into $\mathbf{R}^n$.

Exercise 7.

Show that (X, d) embeds into an Euclidean space if and only if D is non-negative definite.