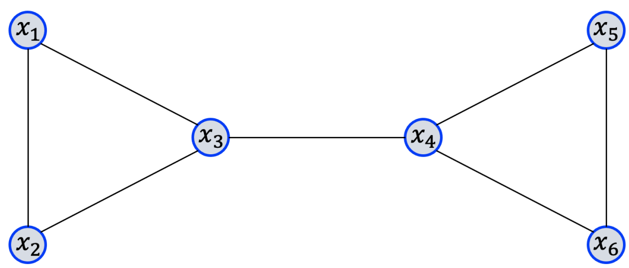
MTH793P **Advanced Machine Learning, Semester B, 2023/24**

Coursework 5

---

1 Evaluating clustering algorithms

In this problem we assume that we have a data set $\{x_1, \dots, x_6\} \subset \mathbb{R}^2$ that looks as follows:



The Euclidean distances along all the edges segments is equal to 1.
We will consider to alternative clusterings on these data:

- $\mathcal{C} = \{C_1, C_2\}$, with $C_1 = \{x_1, x_2, x_3\}$ and $C_2 = \{x_4, x_5, x_6\}$.
- $\mathcal{C}' = \{C'_1, C'_2, C'_3\}$, with $C'_1 = \{x_1, x_2\}$, $C'_2 = \{x_3, x_4\}$, and $C'_3 = \{x_5, x_6\}$.

1. Compute the Dunn-Index (DI) for $\mathcal{C}$ and $\mathcal{C}'$, using the single-linkage inter-cluster distance, and the diameter intra-cluster distance. Which clustering is better?
2. Compute the mean Silhouette Coefficient (SC) for $\mathcal{C}$ and $\mathcal{C}'$. Which one is better?
3. Suppose that we know that $\mathcal{C}'$ is the ground-truth for this dataset. Compute the Rand Index (RI) for $\mathcal{C}$.

Solution:

1. Using the single-linkage distance, we have:

$$\delta(C_1, C_2) = \delta(C'_1, C'_2) = \delta(C'_2, C'_3) = 1, \quad \delta(C'_1, C'_3) = 1 + \sqrt{3}.$$

   Using the diameter:

$$\Delta(C_1) = \Delta(C_2) = \Delta(C'_1) = \Delta(C'_2) = \Delta(C'_3) = 1.$$

Therefore,

$$\mathrm{DI}(\mathcal{C}) = \frac{1}{1} = 1, \quad \mathrm{DI}(\mathcal{C}') = \frac{\min(1, 1+\sqrt{3})}{1} = 1.$$

In other words, the Dunn index we computed does not favour any of the clusterings.

2. Start with $\mathcal{C}$:

$$a(x_1) = a(x_2) = a(x_3) = a(x_4) = a(x_5) = a(x_6) = 1.$$

Next,

$$\begin{aligned} \|x_1 - x_4\| &= \sqrt{(1/2)^2 + (\sqrt{3}/2 + 1)^2} = 1.9319, \\ \|x_1 - x_5\| &= 1 + \sqrt{3} = 2.7321, \\ \|x_1 - x_6\| &= \sqrt{1 + (1+\sqrt{3})^2} = 2.9093. \end{aligned}$$

Therefore,

$$b(x_1) = b(x_2) = b(x_5) = b(x_6) = \frac{1}{3}(1.9319 + 2.7321 + 2.9093) = 2.5244,$$

and

$$b(x_3) = b(x_4) = \frac{1}{3}(1 + 1.9319 + 1.9319) = 1.6213.$$

We conclude that

$$s(x_1) = s(x_2) = s(x_5) = s(x_6) = \frac{2.5244 - 1}{2.5244} = 0.6039,$$

and

$$s(x_3) = s(x_4) = \frac{1.6213 - 1}{1.6213} = 0.3832.$$

Overall, we have

$$\mathrm{SC}(\mathcal{C}) = \frac{1}{6}(4 \times 0.6039 + 2 \times 0.3832) = 0.5303.$$

Next, we do the same for $\mathcal{C}'$:

$$a(x_1) = a(x_2) = a(x_3) = a(x_4) = a(x_5) = a(x_6) = 1.$$

Next,

$$b(x_1) = b(x_2) = b(x_5) = b(x_6) = \frac{1}{2}(1 + 1.9319) = 1.4660,$$

and

$$b(x_3) = b(x_4) = 1.$$

Therefore,

$$s(x_1) = s(x_2) = s(x_5) = s(x_6) = \frac{0.4660}{1.4660} = 0.3179,$$

and

$$s(x_3) = s(x_4) = 0.$$

We conclude that

$$\mathrm{SC}(\mathcal{C}') = \frac{1}{6}(4 \times 0.3179) = 0.2119.$$

For the silhouette coefficient, clearly the $\mathcal{C}$ is better than $\mathcal{C}'$.

3. Since $\mathcal{C}'$ is assumed to be the correct clustering, we have
   - True Positives: $(x_1, x_2), (x_5, x_6)$.
   - True Negatives: $(x_1, x_4), (x_1, x_5), (x_1, x_6), (x_2, x_4), (x_2, x_5), (x_2, x_6), (x_3, x_5), (x_3, x_6)$.

   Therefore

$$\mathrm{RI} = \frac{\mathrm{TP} + \mathrm{TN}}{\binom{6}{2}} = \frac{10}{15} = 0.666.$$

2 SVD

1. Find vectors $u \in \mathbb{R}^2$ and $v \in \mathbb{R}^3$ such that the following identity is satisfied for all known values:

$$uv^\top = \begin{pmatrix} 1 & 0 & ? \\ -2 & ? & 4 \end{pmatrix}.$$

   What value do you obtain at the missing entry denoted by a question mark?

2. Compute the singular value decomposition of the matrix

$$X = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$

   by hand. **Hint**: Find the eigenvalues of $XX^\top$ by computing the characteristic polynomial. Then compute vectors in the nullspace of $X^\top X - \lambda I$, where λ are the roots of the characteristic polynomial and zero, in order to compute the eigenvectors u_1, u_2 and v_1, v_2, v_3 that form the matrices U and V.

3. Compute an approximation $\hat{L} \in \mathbb{R}^{2\times 3}$ with $\mathrm{rank}(\hat{L}) = 1$ of the matrix

$$X := \begin{pmatrix} -2 & 3 & 2 \\ 2 & 2 & 3 \end{pmatrix}$$

   by hand that satisfies $\|\hat{L} - X\|_{\mathrm{Fro}} \leq \|L - X\|_{\mathrm{Fro}}$, for all $L \in \mathbb{R}^{2\times 3}$ with $\mathrm{rank}(L) = 1$.

Solution:

1. In order to satisfy the equality, the 2×3-matrix has to have rank one. Hence, if we choose

$$\begin{pmatrix} 1 & 0 & -2 \\ -2 & 0 & 4 \end{pmatrix}.$$

we ensure that the entries of the first row are the entries of the second row multiplied by -2. This way both rows are linearly dependent, leading to a matrix of rank one. Two possible vectors u and v that satisfy

$$uv^\top = \begin{pmatrix} 1 & 0 & -2 \\ -2 & 0 & 4 \end{pmatrix}$$

are $u = \begin{pmatrix} 1 & -2 \end{pmatrix}^\top$ and $v = \begin{pmatrix} 1 & 0 & -2 \end{pmatrix}^\top$.

2. The main equations to compute SVD are

$$X^\top X = V\Sigma^\top \Sigma V^\top \tag{1}$$

$$XX^\top = U\Sigma\Sigma^\top U^\top \tag{2}$$

Since Σ is diagonal and V is orthogonal, Eq.(1)-(2) show that Σ and V (or U) can be respectively computed from the eigenvalues and the eigenvectors of $X^\top X$ (or $XX^\top$).

If you need to solve this by hand, a **useful trick** is to start with $X^\top X$ if X has more rows than columns, otherwise you should start with $XX^\top$. For this exercise it is better to start with Eq.(2).

$$XX^\top = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}$$

Its eigenvalues can be computed by solving $\det(XX^\top - \lambda I) = 0$.

$$\det(XX^\top - \lambda I) = \det \begin{pmatrix} 17-\lambda & 8 \\ 8 & 17-\lambda \end{pmatrix} = (17-\lambda)^2 - 64 = \lambda^2 - 34\lambda + 17^2 - 64 = 0$$

whose solutions are $\lambda_1 = 25$ and $\lambda_2 = 9$. Since $\sigma_i^2 = \lambda_i$ and $\sigma_i > 0$ for all i, it results $\sigma_1 = 5\,,\ \sigma_2 = 3$.

Now let's compute the eigenvectors of $XX^\top$. It is sufficient to compute the kernel of $XX^\top - \lambda_i I$

$$\bar{u}_1 \in \ker(XX^\top - \lambda_1 I) = \ker \begin{pmatrix} 17-25 & 8 \\ 8 & 17-25 \end{pmatrix} = \ker \begin{pmatrix} -8 & 8 \\ 8 & -8 \end{pmatrix} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} t \right\}$$

$$\bar{u}_2 \in \ker(XX^\top - \lambda_2 I) = \ker \begin{pmatrix} 17-9 & 8 \\ 8 & 17-9 \end{pmatrix} = \ker \begin{pmatrix} 8 & 8 \\ 8 & 8 \end{pmatrix} = \left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix} t \right\}$$

where we can arbitrarily select $t = 1$. $\bar{u}_1$ and $\bar{u}_2$ need to be orthonormal vectors, we have to divide both of them by their modulus:

$$u_1 = \bar{u}_1 \frac{1}{\sqrt{2}}$$

$$u_2 = \bar{u}_2 \frac{1}{\sqrt{2}}$$

u_1 and u_2 are now orthonormal vectors and they are the columns of U accordingly to Eq.(2).

So far we have computed:

$$\Sigma = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}$$

$$U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

To find V, we need to compute the eigenvectors of $X^\top X$:

$$X^\top X = \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} = \begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$$

Its eigenvalues can be computed by solving $\det(X^\top X - \lambda I) = 0$

$$\det(X^\top X - \lambda I) = \det \begin{pmatrix} 13-\lambda & 12 & 2 \\ 12 & 13-\lambda & -2 \\ 2 & -2 & 8-\lambda \end{pmatrix} = 0$$

But we don't have to go through that! Indeed, we already know that, for $XX^\top$, $\lambda_1 = 25$ and $\lambda_2 = 9$. What about λ_3? Since there are only 2 singular values, it must be $\lambda_3 = 0$.

As we did before, compute the kernel of $X^\top X - \lambda_i I$

$$\bar{v}_1 \in \ker(X^\top X - \lambda_1 I) = \ker \begin{pmatrix} 13-25 & 12 & 2 \\ 12 & 13-25 & -2 \\ 2 & -2 & 8-25 \end{pmatrix} = \ker \begin{pmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{pmatrix}$$
$$= \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} t \right\}$$
$$\bar{v}_2 \in \ker(X^\top X - \lambda_2 I) = \ker \begin{pmatrix} 13-9 & 12 & 2 \\ 12 & 13-9 & -2 \\ 2 & -2 & 8-9 \end{pmatrix} = \ker \begin{pmatrix} 4 & 12 & 2 \\ 12 & 4 & -2 \\ 2 & -2 & -1 \end{pmatrix}$$
$$= \left\{ \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix} t \right\}$$
$$\bar{v}_3 \in \ker(X^\top X - \lambda_3 I) = \ker \begin{pmatrix} 13-0 & 12 & 2 \\ 12 & 13-0 & -2 \\ 2 & -2 & 8-0 \end{pmatrix} = \ker \begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$$
$$= \left\{ \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix} t \right\}$$

where we can arbitrarily select $t = 1$. Again, note that $\bar{v}_1$, $\bar{v}_2$, and $\bar{v}_3$ needs to be orthonormal vectors, so we have to divide each of them by their modulus:

$$v_1 = \bar{v}_1 \frac{1}{\sqrt{2}}$$

$$v_2 = \bar{v}_2 \frac{1}{\sqrt{18}}$$

$$v_3 = \bar{v}_3 \frac{1}{\sqrt{9}}$$

v_1, v_2, and v_3 are now orthonormal vectors and they are the columns of V accordingly to Eq.(1). V can then be written as:

$$V = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \\ 1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & -1/3 \end{pmatrix}$$

3. From the lecture notes we know that the best possible rank-one approximation in terms of the Frobenius norm can be computed by computing the (incomplete) singular value decomposition of X. Hence, we compute the eigenvalues of $XX^\top$ by

solving the characteristic polynomial $\det(XX^\top - \lambda I) = 0$, i.e.

$$\det(XX^\top - \lambda I) = \det\left(\begin{pmatrix} 17-\lambda & 8 \\ 8 & 17-\lambda \end{pmatrix}\right) = \lambda^2 - 34\lambda + 225\,,$$

whose solutions are $\lambda_1 = 25$ and $\lambda_2 = 9$. Since the singular values are $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, 2$, we obtain $\sigma_1 = 5$ and $\sigma_2 = 3$. The best rank one approximation can be computed by computing $\tilde{X} = u_1 u_1^\top X$, where u_1 is the singular vector that corresponds to σ_1. We determine u_1 by computing the kernel of $XX^\top - \lambda_1 I$, i.e.

$$\ker(XX^\top - \lambda_1 I) = \ker\left(\begin{pmatrix} -8 & 8 \\ 8 & -8 \end{pmatrix}\right) = \left\{ t \begin{pmatrix} 1 \\ 1 \end{pmatrix} \middle| \, t \in \mathbb{R} \right\}.$$

Since $u_1 \in \ker(XX^\top - \lambda_1 I)$ has to have norm one, we easily compute

$$u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

As a consequence, the best rank-one approximation of X in terms of the Frobenius norm is computed via

$$\begin{aligned} \tilde{X} = u_1 u_1^\top X &= \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} -2 & 3 & 2 \\ 2 & 2 & 3 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 0 & 5 & 5 \\ 0 & 5 & 5 \end{pmatrix}. \end{aligned}$$