Mathematics of Learning – Worksheet 9 - – Discussion on Dec. 14/15th, 2023

- The exercise sheets will be uploaded every Monday. Solution sketches will be uploaded one week later.
- You can hand in your own solutions via StudOn and we correct them - this is not mandatory. Please hand in small groups of 2-3 students.
- For questions, please use the forum on StudOn since other students may have similar questions. If you have a more personal question about the exercises please send an email to ehsan.waiezi@fau.de or lars.weidner@fau.de respectively.

This sheet is easier than the others, it serves as a little motivation to scan over the previous lectures and remember what the key points in these have been. Nevertheless, to prepare for the exam, it is also wise to solve previous sheets if not yet done so. This is ***not*** *a mock exam.*

Exercise 1 [K-means].
Consider the data set $X := \{x_1, ..., x_6\} := \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix}, \begin{pmatrix} -2.5 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ -2 \end{pmatrix} \right\}$, clustered in Clusters $C_1 := \{1, 2, 3\}$ and $C_2 := \{4, 5, 6\}$, with cluster means $m_1 = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix}$ and $m_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Calculate three iterations of the k-means algorithm, starting with the given data.

Solution: Initial state: $C_1 := \{1, 2, 3\}$ and $C_2 := \{4, 5, 6\}$
The cluster means, as given in the exercise, are the arithmetic means:

$$m_1 := \frac{1}{|C_1|} \sum_{i \in C_1} x_i = \frac{1}{3} \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix} + \begin{pmatrix} -2.5 \\ 0 \end{pmatrix} \right) = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix}$$

and furthermore

$$m_2 := \frac{1}{|C_2|} \sum_{i \in C_2} x_i = \frac{1}{3} \left(\begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix} + \begin{pmatrix} -1 \\ -2 \end{pmatrix} \right) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

First iteration:

Assignment step: We assign each data point to the cluster whose mean is the closest of all means to the data point, i.e., x is assigned to cluster number $\operatorname{argmin}_k \|x - m_k\|$.
Since the first component of m_1 and m_2 are equal, the argmin statement reduces to "is the second component of x closer to 0.5 or 1". Hence, we get

$$C_1 := \{2, 3, 6\} \text{ and } C_2 := \{1, 4, 5\}.$$

We update the cluster means, i.e.,

$$m_1 := \frac{1}{|C_1|} \sum_{i \in C_1} x_i = \frac{1}{3} \left(\begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix} + \begin{pmatrix} -2.5 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ -2 \end{pmatrix} \right) = \begin{pmatrix} -\frac{2}{3} \\ -\frac{1}{2} \end{pmatrix}$$

and furthermore

$$m_2 := \frac{1}{|C_2|}\sum_{i\in C_2} x_i = \frac{1}{3}\left(\begin{pmatrix}1\\1\end{pmatrix}+\begin{pmatrix}1\\3\end{pmatrix}+\begin{pmatrix}0\\2\end{pmatrix}\right) = \begin{pmatrix}\frac{2}{3}\\2\end{pmatrix}$$

First iteration is complete.

Second iteration: We again assign each data point to the cluster whose mean is closest to the data points amongst all means, i.e.,

$$C_1 := \{3,6\} \text{ and } C_2 := \{1,2,4,5\}.$$

Update step:

$$m_1 := \frac{1}{|C_1|}\sum_{i\in C_1} x_i = \frac{1}{2}\begin{pmatrix}-\frac{7}{2}\\-2\end{pmatrix} = \begin{pmatrix}-\frac{7}{4}\\-1\end{pmatrix}$$

and

$$m_2 := \frac{1}{|C_2|}\sum_{i\in C_2} x_i = \frac{1}{4}\begin{pmatrix}\frac{5}{2}\\\frac{5}{2}\end{pmatrix} = \frac{1}{8}\begin{pmatrix}7\\13\end{pmatrix}$$

Note that in the next assignment step, x gets assigned to cluster number $\operatorname{argmin}_k \|x - m_k\|$.

$$C_1 := \{3,6\} \text{ and } C_2 := \{1,2,4,5\}.$$

Hence this is the final iteration (since we got the same clusters than last time).

Exercise 2 [Linear PCA].

Given a set of data points $X := \{x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}\} := \left\{\begin{pmatrix}7\\-3.5\\1.5\end{pmatrix}, \begin{pmatrix}4\\-3\\2.5\end{pmatrix}, \begin{pmatrix}9\\-5\\3\end{pmatrix}, \begin{pmatrix}7\\-2\\1\end{pmatrix}\right\}$.

Calculate the first two (linear) principal components of the data set.
Solution: Compute mean value data:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x^{(i)} = \begin{pmatrix}6.75\\-3.375\\2\end{pmatrix}$$

Center data via

$$y^{(i)} = x^{(i)} - \overline{X} \Rightarrow Y = \begin{pmatrix}0.25 & -2.75 & 2.25 & 0.25\\-0.125 & 0.375 & -1.625 & 1.375\\-0.5 & 0.5 & 1 & -1\end{pmatrix}$$

Compute covariance matrix:

$$C = \frac{1}{N}\sum_{i=1}^{N} y^{(i)} y^{(i)^T} = \begin{pmatrix}\frac{51}{16} & -\frac{35}{32} & \frac{1}{8}\\-\frac{35}{32} & \frac{75}{64} & -\frac{11}{16}\\\frac{1}{8} & -\frac{11}{16} & \frac{5}{8}\end{pmatrix}$$

Determine the eigenvalues and eigenvectors of C:
Eigenvalues: $\lambda_1 = 3.718, \lambda_2 = 1,211, \lambda_3 = 0.056$
Eigenvectors:

$$(v_1, v_2, v_3) = \begin{pmatrix} 6.917 & -0.594 & 0.266 \\ -3.241 & -0.960 & 0.877 \\ 1 & 1 & 1 \end{pmatrix}$$

The first two (since $k = 2$) define our transformation matrix T:

$$T := (v_1, v_2) = \begin{pmatrix} 6.917 & -0.594 \\ -3.241 & -0.960 \\ 1 & 1 \end{pmatrix}$$

Remark: This eigenvector matrix works, even though it is somehow smart to take normed eigenvectors, at least if you are interested in transforming the compressed data back to the original data space: If you have normed eigenvectors, you just take T to transform back; if you do not have normed eigenvectors, a few (simple, but nevertheless unnecessarily complicated) transformations for T have to be done. We compute the transformed data points for each centered input point $y^{(i)}$ with respect to the two largest eigenvalues via:

$$T^T y^{(i)} = z^{(i)},$$

hence, the transformed data points are

$$z^{(1)} = \begin{pmatrix} 6.917 & -3.241 & 1 \\ -0.594 & -0.960 & 1 \end{pmatrix} \begin{pmatrix} 0.25 \\ -0.125 \\ -0.5 \end{pmatrix} = \begin{pmatrix} 1.634 \\ -0.529 \end{pmatrix}$$

$$z^{(2)} = \begin{pmatrix} 6.917 & -3.241 & 1 \\ -0.594 & -0.960 & 1 \end{pmatrix} \begin{pmatrix} -2.75 \\ 0.375 \\ 0.5 \end{pmatrix} = \begin{pmatrix} -19.737 \\ 1.774 \end{pmatrix}$$

$$z^{(3)} = \begin{pmatrix} 6.917 & -3.241 & 1 \\ -0.594 & -0.960 & 1 \end{pmatrix} \begin{pmatrix} 2.25 \\ -1.625 \\ 1 \end{pmatrix} = \begin{pmatrix} 21.8310 \\ 1.2226 \end{pmatrix}$$

$$z^{(4)} = \begin{pmatrix} 6.917 & -3.241 & 1 \\ -0.594 & -0.960 & 1 \end{pmatrix} \begin{pmatrix} 0.25 \\ 1.375 \\ -1 \end{pmatrix} = \begin{pmatrix} -3.7276 \\ -2.4687 \end{pmatrix}.$$

Exercise 3 [Kernel methods].

Let data points $X := \{x^{(1)}, x^{(2)}, x^{(3)}\} := \left\{ \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 2 \\ 1 \end{pmatrix} \right\} \subset \mathbb{R}^4$ be given.

a) Specify the polynomial kernel depending on parameters $a \in \mathbb{R}$ and $d \in \mathbb{N}$ and the gaussian kernel depending on a parameter $\sigma > 0$.

Solution: Polynomial kernel: K_p with $K_{ij} = k(x^{(i)}, x^{(j)}) = \left(x^{(i)T} x^{(j)} + a\right)^d$

Gaussian Kernel: K_g with $K_{ij} = k(x^{(i)}, x^{(j)}) = \exp\left(-\frac{\|x^{(i)} - x^{(j)}\|^2}{2\sigma^2}\right)$

b) Calculate the kernel matrix of X for a polynomial kernel with $a = 1$, $d = 2$ and a gaussian kernel with $\sigma = 0.5$.

Solution: for $a = 1, d = 2$ is $K_P = \begin{pmatrix} 361 & 4 & 484 \\ 4 & 100 & 9 \\ 484 & 9 & 961 \end{pmatrix}$

for $\sigma = 0.5$ is $K_g = \begin{pmatrix} 1 & \exp(-50) & \exp(-12) \\ \exp(-50) & 1 & \exp(-94) \\ \exp(-12) & \exp(-94) & 1 \end{pmatrix}$

c) Apply the kernel PCA to the data points and calculate the 2-dimensional kernel principal components of the data set X for a polynomial kernel with $a = 1$ and $d = 2$.

Solution: Use the polynomial kernel matrix K_p here with : $K = \begin{pmatrix} 361 & 4 & 484 \\ 4 & 100 & 9 \\ 484 & 9 & 961 \end{pmatrix}$

We center the kernel matrix and the transformed input data via:

$$\tilde{K} = (K - 1_N K - K 1_N + 1_N K 1_N),$$

whereby

$$1_N := \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \text{i.e., } 1_N \in \mathbb{R}^{N \times N}, (1_N)_{ij} := \frac{1}{N}$$

Remark: What does this matrix do if we multiply it to the K-matrix? Well: multiply 1_N from left: get a matrix whose columns are the average of the corresponding column in K; from right: get a matrix whose rows are the average of the corresponding row in K; from left and from right: get a matrix whose entries are the average of the entries of K. i.e.,

$$\tilde{K} = \frac{1}{9} \begin{pmatrix} 571 & -434 & -137 \\ -434 & 2638 & -2204 \\ -137 & -2204 & 2341 \end{pmatrix}$$

Determine the eigenvalues and eigenvectors of $\tilde{K}$:
Eigenvalues: $\lambda_1 = 523.41, \lambda_2 = 93.259, \lambda_3 = 0$
Eigenvectors (with norm 1):

$$\vec{\alpha} = \begin{pmatrix} 0.054368 & -0.81468 & -0.57735 \\ -0.73272 & 0.36026 & -0.57735 \\ 0.67835 & 0.45443 & -0.57735 \end{pmatrix}$$

or, in vector notation,

$$\vec{\alpha}^{(1)} = \begin{pmatrix} 0.054368 \\ -0.73272 \\ 0.67835 \end{pmatrix}, \vec{\alpha}^{(2)} = \begin{pmatrix} -0.81468 \\ 0.36026 \\ 0.45443 \end{pmatrix}, \vec{\alpha}^{(3)} = \begin{pmatrix} -0.57735 \\ -0.57735 \\ -0.57735 \end{pmatrix}.$$

Remark: Take care, the notation here is a bit complicated. We have $\vec{\alpha}^{(j)}$ is the j-th eigenvector of the centered K-matrix.

Take the 2 eigenvectors of the 2 largest non-vanishing eigenvalues and normalize computed eigenvectors via:

$$\vec{\alpha}^{(j)} = \frac{1}{\sqrt{\lambda_j N}} \vec{\alpha}^{(j)}$$

leads to

$$\vec{\alpha}^{(1)} = \begin{pmatrix} 0.001372 \\ -0.018491 \\ 0.017119 \end{pmatrix}, \vec{\alpha}^{(2)} = \begin{pmatrix} -0.048706 \\ 0.021538 \\ 0.027168 \end{pmatrix}.$$

We forget about the third eigenvector, since we want to transform to 2 dimensions; in matrix notation, we get then

$$\vec{\alpha} = \begin{pmatrix} 0.001372 & -0.048706 \\ -0.018491 & 0.021538 \\ 0.017119 & 0.027168 \end{pmatrix}$$

Remark: The notation gets more complicated. $\vec{\alpha}_i^{(j)}$ is the i-th entry of the j-th eigenvector. So, lets have a look at the formula from lecture script. We want to reduce the dimension of an arbitrary data point x in the original data space to a vector z in the reduced dimension data space, then we can do it like this:

$$z_j = \sum_{i=1}^{N} \vec{\alpha}_i^{(j)} k(x^{(i)}, x), \text{ for } \quad j = 1, ..., k$$

Remark: k has two meanings; first: our kernel function; second: the dimension we want to transform our data to. Which is meant depends on context. z_j is the j-th component of the vector z.

Basically, in our case with 3 data vectors,

$$z_j = \vec{\alpha}^{(j)T} \cdot \begin{pmatrix} k(x^{(1)}, x) \\ k(x^{(2)}, x) \\ k(x^{(3)}, x) \end{pmatrix}, \text{ for } \quad j = 1, ..., k$$

or in even more compact notation,

$$z = \vec{\alpha}^T \cdot \begin{pmatrix} k(x^{(1)}, x) \\ k(x^{(2)}, x) \\ k(x^{(3)}, x) \end{pmatrix}.$$

In case we are interested in the transformed versions of not any data points, but the initial data points, we can recycle our kernel matrix, since we calculated values for the kernel function of original data point pairs already. Notation: $z^{(1)}$ is the transformed version of the first data point $x^{(1)}$ and so on. $K_{:,j}$ is denoting the j-th column of K (the kernel matrix, not the centered kernel matrix).

$$z^{(1)} = \vec{\alpha}^T \cdot K_{:,1} = \begin{pmatrix} 8.7069 \\ -4.3474 \end{pmatrix}, z^{(2)} = \vec{\alpha}^T \cdot K_{:,2} = \begin{pmatrix} -1.6895 \\ 2.2035 \end{pmatrix}, z^{(3)} = \vec{\alpha}^T \cdot K_{:,3} = \begin{pmatrix} 16.949 \\ 2.7285 \end{pmatrix}.$$