Solutions worksheet 01

Maksimov, Dmitrii
dmitrii.maksimov@fau.de

Ilia, Dudnik
ilia.dudnik@fau.de

Isa, Baghirov
isa.baghirov@fau.de

Yulia, Herasymenko
yuliia.herasymenko@fau.de

October 28, 2021

Exercise [Solving linear equation systems]

Solve the linear equation system $Ax = b$. A and b given as

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 8 & 27 \end{pmatrix}, b = \begin{pmatrix} 3 \\ 2 \\ 7 \end{pmatrix}$$

Gauss elimination method:

1. Augmented matrix
$$\begin{bmatrix} 1 & 2 & 3 & | & 3 \\ 1 & 4 & 9 & | & 2 \\ 1 & 8 & 27 & | & 7 \end{bmatrix}$$

2. Add -1 times row (1) to both row (2) and row (3)
$$\begin{bmatrix} 1 & 2 & 3 & | & 3 \\ 0 & 2 & 6 & | & -1 \\ 0 & 6 & 24 & | & 4 \end{bmatrix} \begin{matrix} \\ R_2 - R_1 \\ R_3 - R_1 \end{matrix}$$

3. Divide all terms in row (2) by 2 and add -3 times row (2) to row (3)
$$\begin{bmatrix} 1 & 2 & 3 & | & 3 \\ 0 & 1 & 3 & | & -\frac{1}{2} \\ 0 & 0 & 6 & | & 7 \end{bmatrix} \begin{matrix} \\ \frac{R_2}{2} \\ R_3 - 3R_2 \end{matrix}$$

4. Divide all terms in row (3) by 6
$$\begin{bmatrix} 1 & 2 & 3 & | & 3 \\ 0 & 1 & 3 & | & -\frac{1}{2} \\ 0 & 0 & 1 & | & \frac{7}{6} \end{bmatrix} \begin{matrix} \\ \\ \frac{R_3}{6} \end{matrix}$$

5. Row echelon form

   Add -3 times row (3) to both row (2) and row (1)

$$\begin{array}{c} \left[\begin{array}{ccc|c} 1 & 2 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & \frac{7}{6} \end{array}\right] \end{array}\begin{array}{l} R_1 - 3R_3 \\ R_2 - 3R_3 \\ \end{array}$$

   Add -2 times row (2) to row (1)

$$\begin{array}{c} \left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{15}{2} \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & \frac{7}{6} \end{array}\right] \end{array}\begin{array}{l} R_1 - 2R_2 \\ \\ \end{array}$$

Answer:

$$x = \begin{pmatrix} \frac{15}{2} \\ -4 \\ \frac{7}{6} \end{pmatrix}$$

Exercise [Norms]

Definition (Norm) — A mapping $\|\cdot\|$ from any (real) vector space V to the real numbers $\mathbb{R}$ is called a norm, whenever

1. $\|v + w\| \leq \|v\| + \|w\|$
2. $\|v\| = 0 \implies v = 0_V$
3. $\|\lambda v\| = |\lambda| \cdot \|v\|$

for all $\lambda \in \mathbb{R}$, $v, w \in V$

1. Let$V = \mathbb{R}^n$ for some $n \in \mathbb{N}$. The euclidean norm

$$\|v\|_2 := \sqrt{\sum_{i=1}^{n} v_i^2}$$

   is a norm.

   Proof:

   (a) Proof of 1 statement

$$\|v + w\|_2^2 := \sum_{i=1}^{n}(v_i+w_i)^2 = \sum_{i=1}^{n} v_i^2+2v_iw_i+w_i^2 = \sum_{i=1}^{n} v_i^2+\sum_{i=1}^{n} 2v_iw_i+\sum_{i=1}^{n} w_i^2 = \|v\|_2^2+2(v\cdot w)+\|w\|_2^2$$

   Taking into account the Cauchy-Schwarz Inequality

$$|v \cdot w| \leq \|v\|_2 \cdot \|w\|_2$$

   which implies

$$\|v\|_2^2 + 2(v \cdot w) + \|w\|_2^2 \leq \|v\|_2^2 + 2\|v\|\|w\| + \|w\|_2^2 = (\|v\| + \|w\|)^2$$

Hence,

$$\|v + w\|_2^2 \le (\|v\|_2 + \|w\|_2)^2$$
$$\|v + w\|_2 \le \|v\|_2 + \|w\|_2$$

as required

(b) Proof of 2 statement

$$\|v\|_2 := \sqrt{\sum_{i=1}^{n} (v_i)^2}$$
$$\sqrt{\sum_{i=1}^{n} v_i^2} = 0 \iff v_i = 0, \ \forall i$$

which implies

$$v = 0_V$$

as required

(c) Proof of 3 statement

$$\|\lambda v\|_2 := \sqrt{\sum_{i=1}^{n} (\lambda v_i)^2} = |\lambda| \cdot \sqrt{\sum_{i=1}^{n} v_i^2} = |\lambda| \cdot \|v\|$$

as required

Hence, the euclidean norm is a norm.

2. Let$V = \mathbb{R}^n$ for some $n \in \mathbb{N}$. The mapping

$$\|v\|_{\frac{1}{2}} := (\sum_{i=1}^{n} \sqrt{|v_i|})^2$$

is a norm.

Proof: let $v = (0, 1)$, $w = (1, 0)$, so $v, w \in V$. Given that

$$\|v + w\|_{\frac{1}{2}} := (\sum_{i=1}^{2} \sqrt{|v_i + w_i|})^2 = 2^2 = 4$$

$$\|v\|_{\frac{1}{2}} + \|w\|_{\frac{1}{2}} := (\sum_{i=1}^{2} \sqrt{|v_i|})^2 + (\sum_{i=1}^{2} \sqrt{|w_i|})^2 = 1 + 1 = 2$$
$$\|v + w\|_{\frac{1}{2}} > \|v\|_{\frac{1}{2}} + \|w\|_{\frac{1}{2}}$$

Hence, the $\|\cdot\|_{\frac{1}{2}}$ is not a norm.

3. Let V be the space of convergent sequences. The mapping

$$\|v\|_{lim} := \lim_{n \to \infty} v_n$$

is a norm.

Proof:

$$\|\lambda \cdot v\|_{lim} := \lim_{n \to \infty} \lambda \cdot v_n = \lambda \cdot \lim_{n \to \infty} v_n = \lambda \cdot \|v\|_{lim} \ne |\lambda| \cdot \|v\|_{lim}$$

Hence, the $\|\cdot\|_{lim}$ is not a norm.

Exercise [Python, Pandas, K-Means]

(a) DS analysis

Let's start by describing DS

	eruptions	waiting
count	272.000000	272.000000
mean	3.487783	70.897059
std	1.141371	13.594974
min	1.600000	43.000000
25%	2.162750	58.000000
50%	4.000000	76.000000
75%	4.454250	82.000000
max	5.100000	96.000000

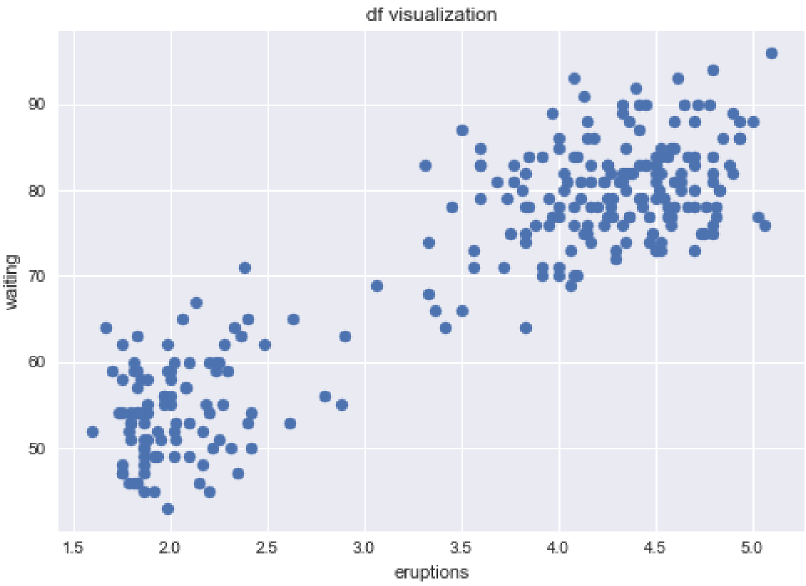
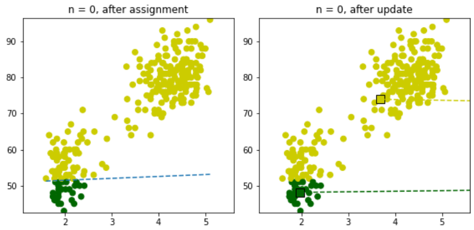
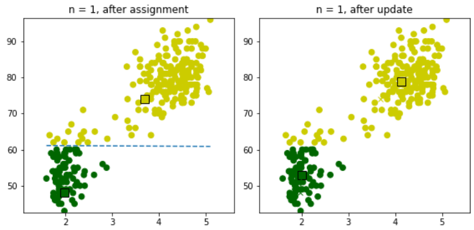


Figure 1: visualization of *faithful.csv* data

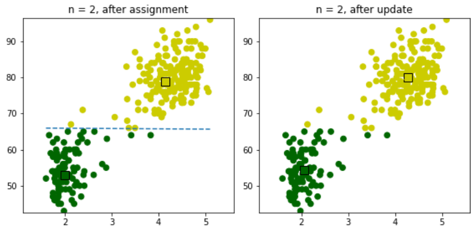
(b) KMeans clustering visualization



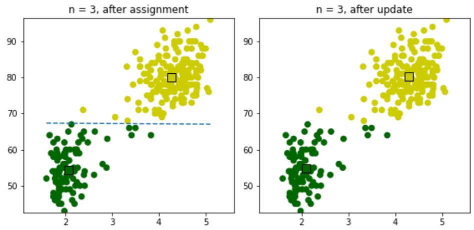
(a) iteration = 0



(b) iteration = 1



(c) iteration = 2



(d) iteration = 3

Figure 2: KMeans algorithm

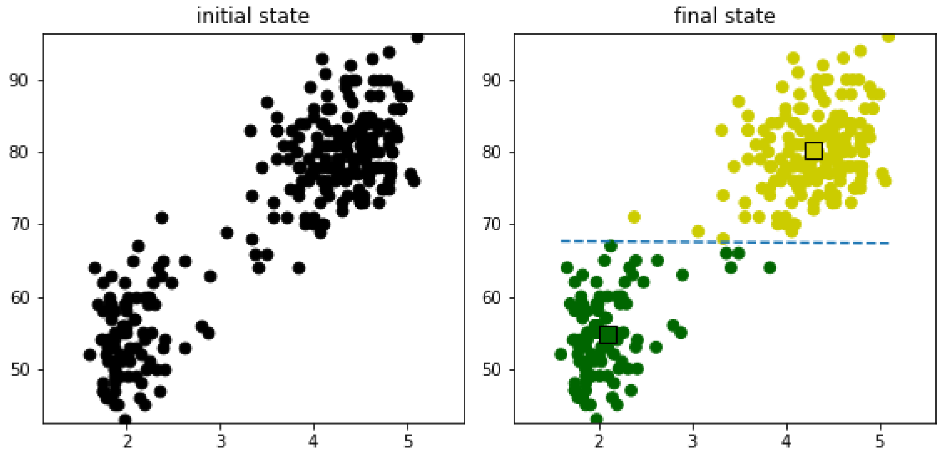
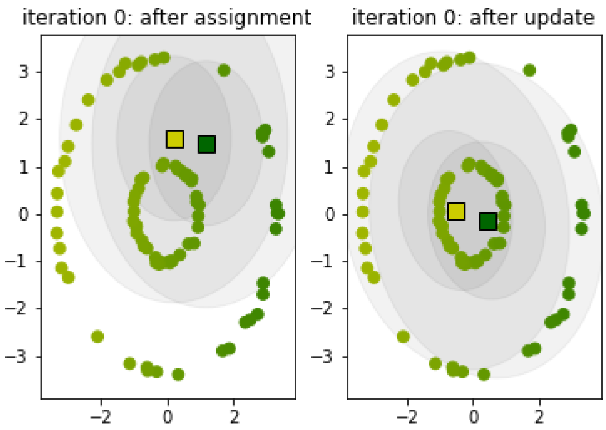


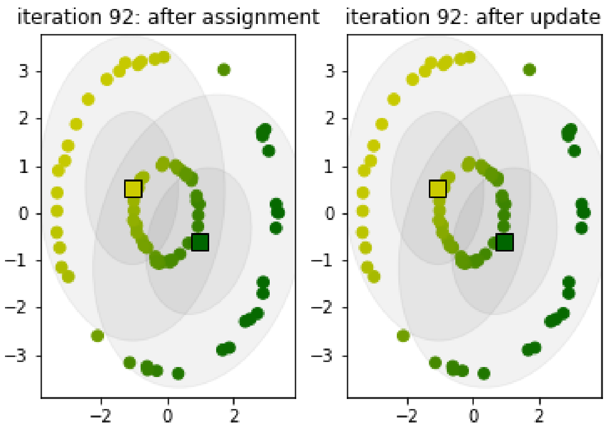
Figure 3: KMean clustering sumup

Exercise [Implementing EM for Clustering]

Since the result of the EM algorithm is predicatable for the *faithful* DS we will look at another dataset.



(a) start



(b) end

Figure 4: EM algorithm

It cannot be denied that result is insufficient. **Hence, we have a question: why is the clustering result like this? It's obvious that there are 2 clusters which can be modulated as random points from multivariate Gaussian distributions with the same means and different covariance matrices.**

Exercise [Experiments with K-Means and EM]

Create a DS of 3 clusters:

(a) apple

(b) desk

(c) classroom

Figure 5: examples of dataset

Let's transform to gray-scale matrices and apply TSNE algorithm to visualize clusters. Plot 2d space to show data:

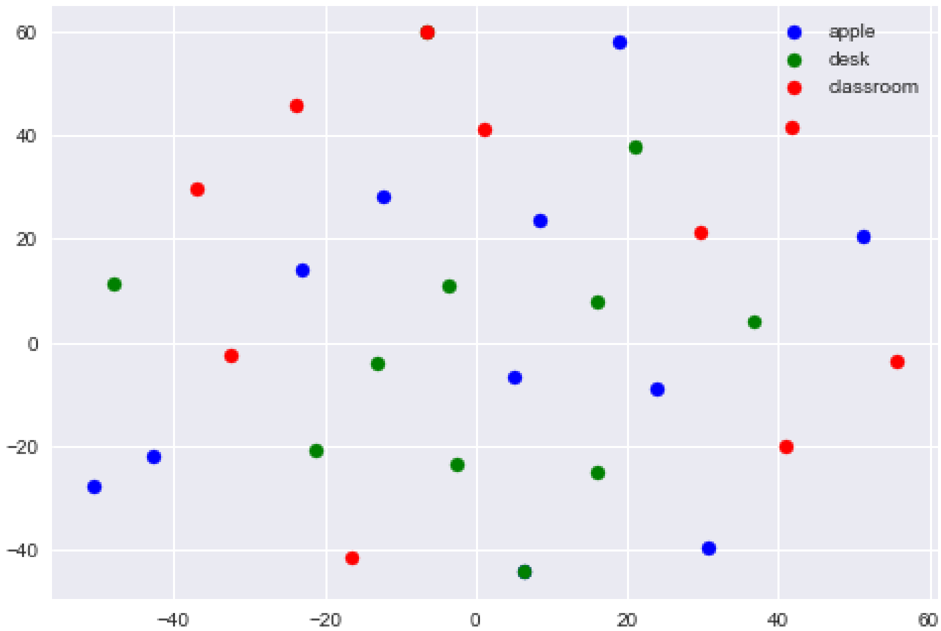


Figure 6: apples, desks, classroomes data

There is no any structure in this data. Hence, there is no point to do clustering. **What did we do wrong?**

Exercise [Theory of K-means]

- Show that the iterates of the algorithm satisfy

$$\frac{1}{2}\sum_{k=1}^{K}\sum_{x\in C_k^{(i)}}\left\|x - m_k^{(i)}\right\|^2 \le \frac{1}{2}\sum_{k=1}^{K}\sum_{x\in C_k^{(i-1)}}\left\|x - m_k^{(i-1)}\right\|^2$$

  To confirm the inequality above we will show that both of operations can never increase the clustering energy.

$$E(C^{(i)}, m^{(i)}) := \frac{1}{2}\sum_{k=1}^{K}\sum_{x\in C_k^{(i)}}\left\|x - m_k^{(i)}\right\|^2 \triangleq \text{clustering energy}$$

1. $E(C^{(i)}, m^{(i)}) < E(C^{(i-1)}, m^{(i)})$

   From the logic of the algorithm: $C^{(i)}$ and $C^{(i-1)}$ are differently only if there is a point that finds a closer cluster center in $m^{(i)}$ than the one assigned to it by $C^{(i-1)}$.Hence,

$$\frac{1}{2}\sum_{k=1}^{K}\sum_{x\in C_k^{(i)}}\left\|x - m_k^{(i)}\right\|^2 < \frac{1}{2}\sum_{k=1}^{K}\sum_{x\in C_k^{(i-1)}}\left\|x - m_k^{(i)}\right\|^2$$

2. $E(C^{(i)}, m^{(i)}) \le E(C^{(i)}, m^{(i-1)})$

   This statement is equivalent to the following

$$\frac{1}{2}\sum_{k=1}^{K}\sum_{x\in C_k}\|x - m_k\|^2 \le \frac{1}{2}\sum_{k=1}^{K}\sum_{x\in C_k}\|x - a_k\|^2$$

   , where $a = (a_1, \dots, a_k)$ with $a_k \in \mathbb{R}^M$ is an arbitrary point in the same space. Consider C_k for it:

$$\sum_{x\in C_k}\|x - m_k\|^2 \le \sum_{x\in C_k}\|x - a_k\|^2$$

$$\begin{aligned}
\sum_{x\in C_k}\|x - a_k\|^2 &= \sum_{x\in C_k}\|(x - m_k) + (x - a_k)\|^2 \\
&= \sum_{x\in C_k}\|x - m_k\|^2 + \|m_k - a_k\|^2 + 2(x - m_k)\cdot(m_k - a_k) \\
&= \sum_{x\in C_k}\|x - m_k\|^2 + \sum_{x\in C_k}\|m_k - a_k\|^2 + 2\sum_{x\in C_k}(x\cdot m_k - x\cdot a_k - m_k\cdot m_k + m_k\cdot a_k) \\
\text{as} \sum_{x\in C_k} x &= \sum_{x\in C_k} m_k \\
&= \sum_{x\in C_k}\|x - m_k\|^2 + |C_k|\|m_k - a_k\|^2 + 2\cdot|C_k|(m_k\cdot m_k - m_k\cdot a_k - m_k\cdot m_k + m_k\cdot a_k) \\
&= \sum_{x\in C_k}\|x - m_k\|^2 + |C_k|\|m_k - a_k\|^2 \\
&\ge \sum_{x\in C_k}\|x - m_k\|^2
\end{aligned}$$

Hence,

$$\frac{1}{2}\sum_{k=1}^{K}\sum_{x\in C_k^{(i)}}\left\|x-m_k^{(i)}\right\|^2 \leq \frac{1}{2}\sum_{k=1}^{K}\sum_{x\in C_k^{(i-1)}}\left\|x-m_k^{(i-1)}\right\|^2$$

- Every data point x_n must be assigned to precisely one class in oder to the algorithm be converging. It was shown above.
- It is not a big deal to extend first step to an arbitrary norm. The problem is how to calculate centers of clusters.
- In that case result of clustering depends on realization in programm. Final result significantly depends on the initial statements of centers.

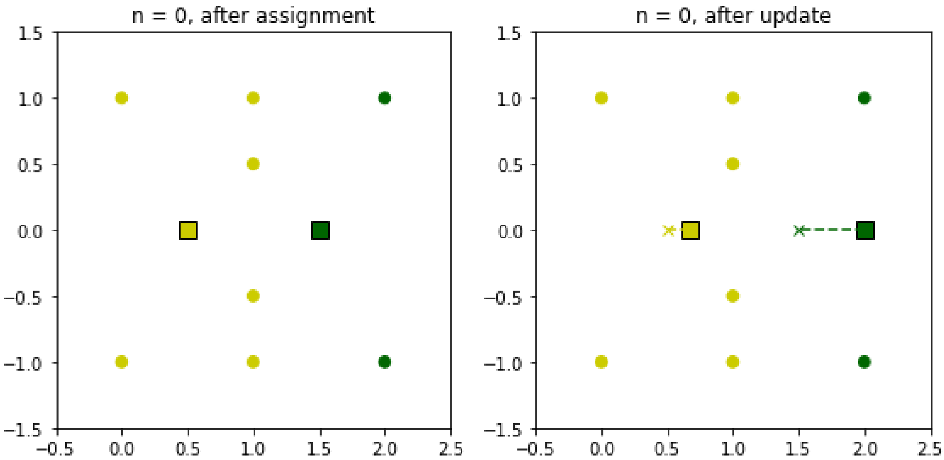


Figure 7: KMeans algorithm

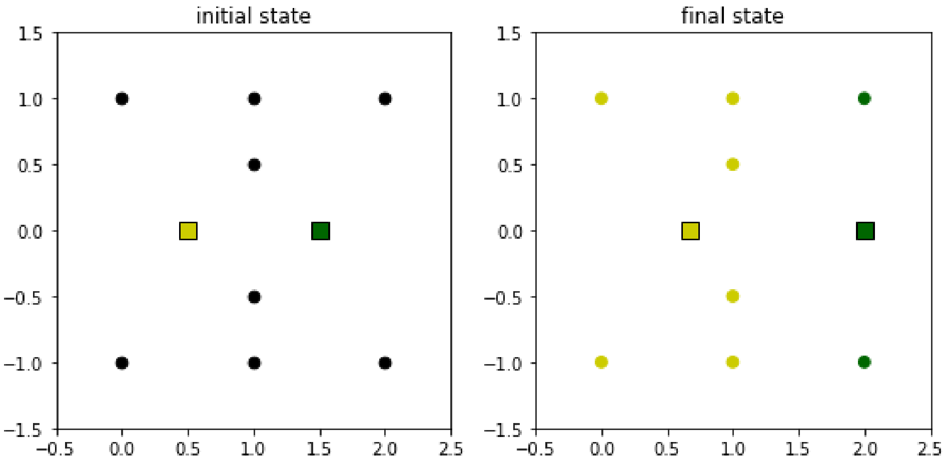


Figure 8: KMean clustering sumup