

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

November 5, 2021

Exercise [Analytically calculating eigenvalues and eigenvectors]

Let $A \in \mathbb{R}^{n \times n}$ be a quadratic matrix. Whenever for a vector $v \in \mathbb{R}^n$ and for a $\lambda \in \mathbb{R}$ the equation

$$Av = \lambda v \quad (1.1)$$

holds, we call λ an eigenvalue of A and v the corresponding eigenvector. Find out how to calculate eigenvalues and eigenvectors analytically and calculate the eigenvalues and eigenvectors of the matrix

$$A := \frac{1}{3} \cdot \begin{pmatrix} 5 & -2 & -3 \\ -2 & 5 & 1 \\ -1 & 1 & 8 \end{pmatrix}.$$

Given 1.1, bring all to left hand side and put in an identity matrix I :

$$v(A - \lambda I) = 0$$

If v is non-zero then:

$$|A - \lambda I| = 0 \quad (1.2)$$

$$\begin{aligned} |A - \lambda I| &:= \left(\frac{1}{3}\right)^3 \cdot \begin{vmatrix} 5 - 3\lambda & -2 & -3 \\ -2 & 5 - 3\lambda & 1 \\ -1 & 1 & 8 - 3\lambda \end{vmatrix} \\ &= \frac{1}{27} (5 - 3\lambda)((5 - 3\lambda)(8 - 3\lambda) - 1) + 2(2(3\lambda - 8) + 1) - (-2 + 5 - 3\lambda) \\ &= -\frac{1}{27} \cdot \frac{1}{27} (\lambda^3 - 6\lambda^2 + 11\lambda - 6) \\ &= -\frac{1}{27} \cdot \frac{1}{27} (\lambda - 1)(\lambda - 2)(\lambda - 3) \end{aligned}$$

Solving 1.2: $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3$.

Let us find their matching eigenvectors using 1.1:

$$\begin{pmatrix} 5 & -2 & -3 \\ -2 & 5 & 1 \\ -1 & 1 & 8 \end{pmatrix} \begin{pmatrix} v_{i_1} \\ v_{i_2} \\ v_{i_3} \end{pmatrix} = \lambda_i \begin{pmatrix} v_{i_1} \\ v_{i_2} \\ v_{i_3} \end{pmatrix}, \forall i \in [1, 2, 3]$$

Hence, $v_1 = (1, 1, 0)^T$, $v_2 = (-1, 1, -1)^T$, $v_3 = (-1, 1, 2)^T$.

Exercise [Definiteness of the covariance matrix]

Let $y^{(1)}, \dots, y^{(N)} \in \mathbb{R}^M$ be centered input data and let C be the respective covariance matrix.

1. Show that C is always positive semi-definite.

If C is positive semi-definite matrix then $x^T C x \geq 0, \forall x \in \mathbb{R}^M$.

$$\begin{aligned} \langle x, Cx \rangle &= \langle x, \left(\frac{1}{N} \sum_{i=1}^N y_i y_i^T \right) x \rangle \\ &= \frac{1}{N} \sum_{i=1}^N x^T y_i y_i^T x \\ &= \frac{1}{N} \sum_{i=1}^N \langle x, y_i \rangle \langle y_i, x \rangle \\ &= \frac{1}{N} \sum_{i=1}^N \langle x, y_i \rangle^2 \geq 0 \end{aligned}$$

2. In which cases is $\langle y, Cy \rangle = 0$ for $y \in \mathbb{R}^M / \{0\}$. What does that mean for the given data?

$$\langle y, Cy \rangle = 0 \iff \text{rank}(C) < N$$

This can be observed in 2 cases:

- $N < M$

It means that observations less than number of features. This is common in image datasets where a feature is a pixel.

- $\text{Var}(\text{feature}) = 0$

It means that there is no dispersion in some feature, i.e. a feature is constant.

Exercise [Implementing PCA for data reduction]

Plot the features after applying the PCA algorithm for $k = 3$ and $k = 2$. What can you observe?

Honestly, we don't understand what we have to do. So, we will do following steps:

1. Plot the features after applying the PCA algorithm for $k = 3$, since the distributions of features don't depend on number of k .
2. Show how much of the variance in data is explained depending on k

Plot the features after applying the PCA algorithm for $k = 3$.

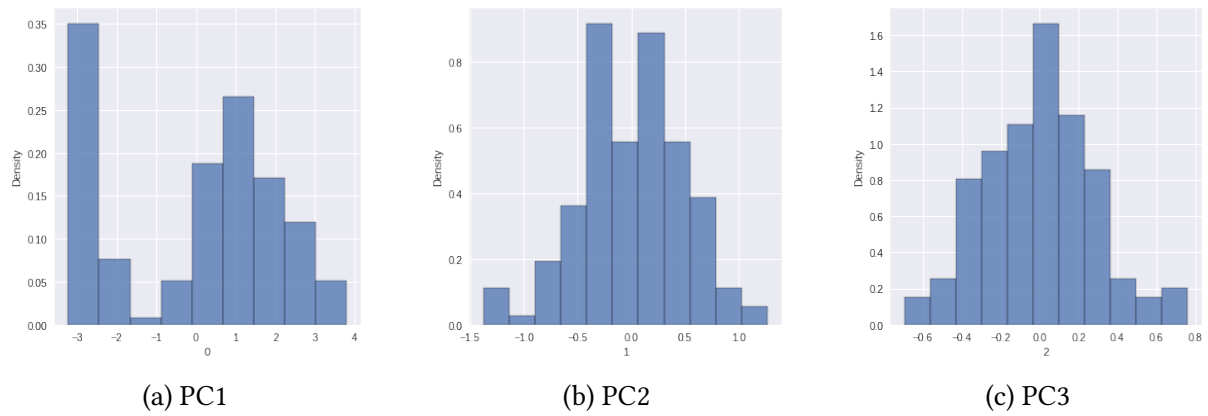


Figure 1: Distribution of PC features

The only thing we can say about these distributions is that they are centered. Now, let us see how much variance is explained depending on number of PC.

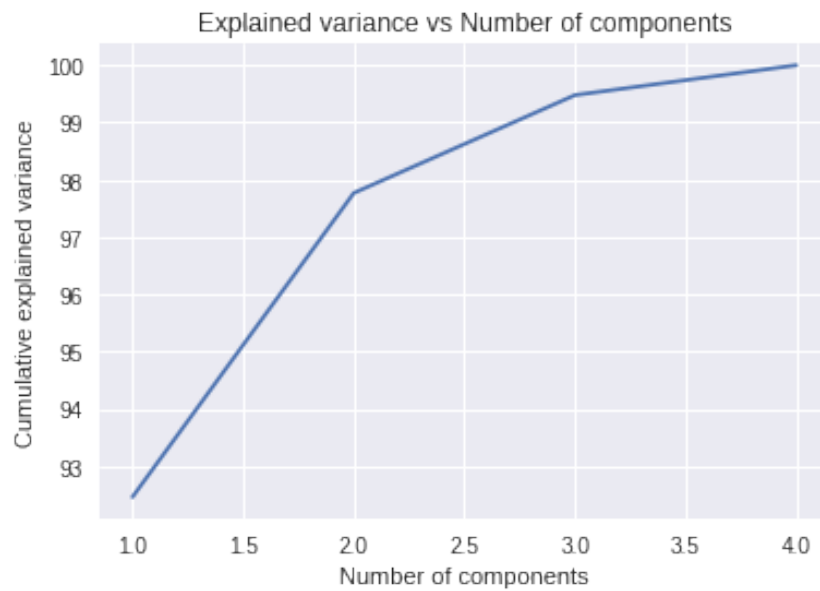


Figure 2: Explained variance

If we use the first feature, it will explain 92% of the data; if we use two features we are able to capture approximately 95% of the data. If we use all features we are going to describe the entire dataset.

Exercise [Apply clustering algorithms]

The code implementation can be founded here [Code](#). In the first part we implement KMeans and EM algorithms on original DS and then we reduce dimension in order to visualize results.

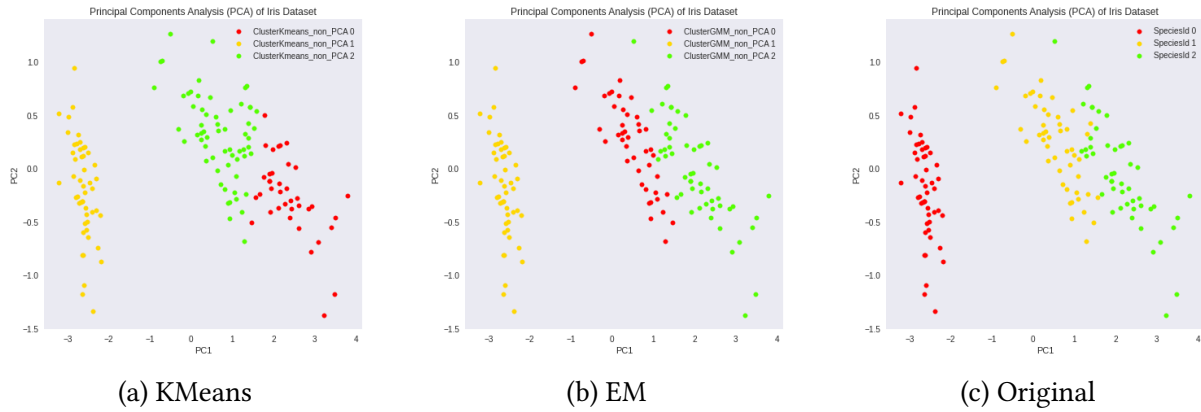


Figure 3: KMeans and EM clustering on original DS

In the second part we implement KMeans and EM algorithms on transformed DS via PCA with 2 components.

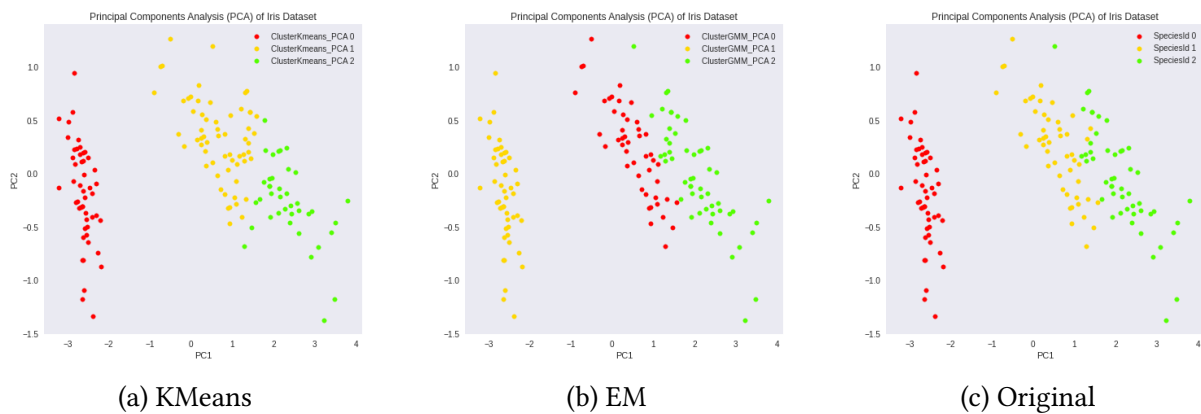


Figure 4: KMeans and EM clustering on transformed DS

Comparing results of EM and KMeans algorithms it is obvious that EM clustering performs better on the given DS. This is because, features are likely to be normally distributed. Hence, the EM clustering worked better at finding the actual species labels. Speaking about PCA implementation for clustering, in this case we use only 2 features, but the result remained approximately the same, since this 2 features explain 95% of original data.