Семинар №7. Задача №2

выполнила студентка группы РК6-56Б, Новичкова Мария

1 Задание

Дано пять наблюдений:

$$
\begin{aligned}
\boldsymbol{x}^{(1)} &= [1.1, 1.3, 1.5, 1.55, 1.6, 1.9, 2, 2.1] \\
\boldsymbol{x}^{(2)} &= [2, 0.9, 0.7, 1.5, 2.6, 0.3, 0.8, 1.4] \\
\boldsymbol{x}^{(3)} &= [-2.9, -0.5, 0.1, -1.5, -3.6, 1.3, 0.4, -0.7] \\
\boldsymbol{x}^{(4)} &= [1.1, 0.2, 0.1, 0.6, 1.3, -0.4, -0.1, 0.4] \\
\boldsymbol{x}^{(5)} &= [0.9, -0.4, -0.8, -0, 1, -1.6, -1.2, -0.7]
\end{aligned}
$$

Требуется найти главные компоненты данных и приблизительно оценить, какова размерность подпространства, из которого были взяты эти наблюдения.

2 Решение

Перед нахождением главные компоненты матрицы необходимо центрировать матрицу. Матрица центрированных данных $\boldsymbol{A}$ получается из матрицы данных $\boldsymbol{X}$ путем следующих преобразований:

$$\boldsymbol{A} = \boldsymbol{X} - \boldsymbol{e}\bar{\boldsymbol{x}} = \left(\boldsymbol{E} - \frac{1}{m}\boldsymbol{e}\boldsymbol{e}^T\right)\boldsymbol{X}$$

Полученная центрированная матрица:

$$\boldsymbol{A} = \begin{bmatrix} 0.66 & 1. & 1.18 & 1.12 & 1.02 & 1.6 & 1.62 & 1.6 \\ 1.56 & 0.6 & 0.38 & 1.07 & 2.02 & 0. & 0.42 & 0.9 \\ -3.34 & -0.8 & -0.22 & -1.93 & -4.18 & 1. & 0.02 & -1.2 \\ 0.66 & -0.1 & -0.22 & 0.17 & 0.72 & -0.7 & -0.48 & -0.1 \\ 0.46 & -0.7 & -1.12 & -0.43 & 0.42 & -1.9 & -1.58 & -1.2 \end{bmatrix}.$$

Главными компонентами матрицы центрированных данных $\boldsymbol{A}$ являются её сингулярные вектора, при этом j-ая главная компонента соответствует j-му сингулярному вектору q_j и стандартному отклонению $\sqrt{\nu}\sigma_j$, где σ_j является j-м сингулярным числом, а коэффициент ν зависит от выбора оценки ковариации (по умолчанию $\nu = \frac{1}{m-1}$).
Сингулярными числами $\sigma_1, \ldots, \sigma_r$ матрицы $\boldsymbol{A} \in R^{(m\times n)}$ называются неотрицательные вещественные числа $\sigma_i = \sqrt{\lambda_i}$, где λ_i являются ненулевые собственные числа соответствующей матрица Грама $\boldsymbol{K} = \boldsymbol{A}^T\boldsymbol{A}$. Ассоциированные собственные вектора матрицы Грама $\boldsymbol{K}$ называются сингулярными векторами.

Матрица Грама для матрицы A:

$$\boldsymbol{K} = \begin{bmatrix} 14.672 & 3.88 & 1.446 & 8.769 & 18.454 & -3.62 & 0.614 & 5.85 \\ 3.88 & 2.5 & 2.39 & 3.59 & 5.21 & 2.2 & 3.01 & 3.95 \\ 1.446 & 2.39 & 2.888 & 2.597 & 2.262 & 3.95 & 3.942 & 3.86 \\ 8.769 & 3.59 & 2.597 & 6.338 & 11.313 & 0.56 & 2.823 & 5.57 \\ 18.454 & 5.21 & 2.262 & 11.313 & 23.288 & -3.85 & 1.408 & 7.89 \\ -3.62 & 2.2 & 3.95 & 0.56 & -3.85 & 7.66 & 5.95 & 3.71 \\ 0.614 & 3.01 & 3.942 & 2.823 & 1.408 & 5.95 & 5.528 & 4.89 \\ 5.85 & 3.95 & 3.86 & 5.57 & 7.89 & 3.71 & 4.89 & 6.26 \end{bmatrix}.$$

Собственные числа и вектора матрицы K (были найдены с помощью функции np.linalg.eig()):

$$\boldsymbol{\lambda} = \begin{bmatrix} 49.07961 & 20.04286 & 0.01088 & 0.00065 & 0. & 0. & 0. & 0. \end{bmatrix}.$$

$$\boldsymbol{P} = \begin{bmatrix} -0.53211 & -0.19665 & 0.22733 & -0.17372 & -0.59105 & -0.0922 & 0.32626 & -0.29511 \\ -0.18056 & 0.21179 & -0.27955 & -0.09946 & 0.58076 & -0.17673 & 0.48266 & -0.47444 \\ -0.10672 & 0.34063 & 0.56412 & -0.06892 & 0.20367 & 0.67322 & 0.4286 & -0.41202 \\ -0.35239 & 0.11006 & -0.0196 & 0.92915 & 0. & -0. & -0. & 0. \\ -0.67818 & -0.18868 & -0.34021 & -0.24203 & 0.19787 & 0.29098 & -0.29372 & 0.12299 \\ 0.04593 & 0.614 & -0.20832 & -0.0597 & -0.40342 & 0.17034 & -0.27799 & -0.26 \\ -0.09925 & 0.5016 & -0.38932 & -0.10527 & -0.18192 & -0.08089 & 0.22338 & 0.32605 \\ -0.27713 & 0.35233 & 0.49077 & -0.13649 & 0.1921 & -0.62194 & -0.51327 & 0.57359 \end{bmatrix}.$$

Тогда главные компоненты матрицы $\boldsymbol{A}$ равны $\boldsymbol{P}^T$:

$$\boldsymbol{M} = \begin{bmatrix} -0.53211 & -0.18056 & -0.10672 & -0.35239 & -0.67818 & 0.04593 & -0.09925 & -0.27713 \\ -0.19665 & 0.21179 & 0.34063 & 0.11006 & -0.18868 & 0.614 & 0.5016 & 0.35233 \\ 0.22733 & -0.27955 & 0.56412 & -0.0196 & -0.34021 & -0.20832 & -0.38932 & 0.49077 \\ -0.17372 & -0.09946 & -0.06892 & 0.92915 & -0.24203 & -0.0597 & -0.10527 & -0.13649 \\ -0.59105 & 0.58076 & 0.20367 & 0. & 0.19787 & -0.40342 & -0.18192 & 0.1921 \\ -0.0922 & -0.17673 & 0.67322 & -0. & 0.29098 & 0.17034 & -0.08089 & -0.62194 \\ 0.32626 & 0.48266 & 0.4286 & -0. & -0.29372 & -0.27799 & 0.22338 & -0.51327 \\ -0.29511 & -0.47444 & -0.41202 & 0. & 0.12299 & -0.26 & 0.32605 & 0.57359 \end{bmatrix}.$$

Далее были рассчитаны соответствующие стандартные отклонения для главных векторов по формуле:

$$\sqrt{\nu}\sigma_j = \sqrt{\frac{1}{m-1}\lambda_i}.$$

И был построен график отклонений относительно соответствующих компонент.

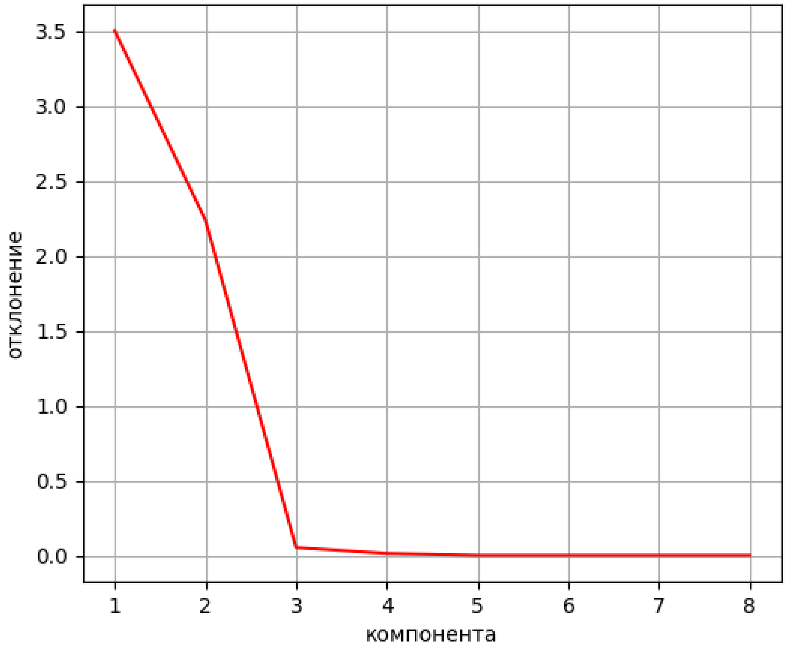

Рис. 1: График отклонений относительно соответствующих компонент

Можно заметить что уже для третьей компоненты отклонение очень близко к нулю, следовательно размерность подпространства, из которого были взяты наблюдения равна 2.
Таким образом можем оставить только первые 2 главные компоненты, так как они содержат большее количество информации.
Тогда главные компоненты:

$$\boldsymbol{M} = \begin{bmatrix} -0.53211 & -0.18056 & -0.10672 & -0.35239 & -0.67818 & 0.04593 & -0.09925 & -0.27713 \\ -0.19665 & 0.21179 & 0.34063 & 0.11006 & -0.18868 & 0.614 & 0.5016 & 0.35233 \end{bmatrix}$$

Таким образом, исходные данные могут быть преобразованы с помощью выбранных компонентов. Размерность данных будет уменьшена до 2.

$$\boldsymbol{A}_{new} = \boldsymbol{A}\boldsymbol{M}^T = \begin{bmatrix} -2.27481 & 2.77347 \\ -3.01706 & 0.21413 \\ 5.8366 & 1.18996 \\ -0.8147 & -1.04885 \\ 0.26995 & -3.12871 \end{bmatrix}.$$