

Homework 1 Undergraduate

September 15, 2021

Jose Carlos Munoz

For Part 1, the math is done on the file HandMadePCA.R

Part1)

1)

It is given that the Data set is set up as D.

$$D = \begin{bmatrix} 19 & 12 \\ 22 & 6 \\ 6 & 9 \\ 3 & 15 \\ 2 & 13 \\ 20 & 5 \end{bmatrix}$$

To center it, μ must be found for both attributes and subtracted from D.

$$\mu_1 = (19 + 22 + 6 + 3 + 2 + 10)/60$$

$$= 12$$

$$\mu_2 = (12 + 6 + 9 + 15 + 13 + 5)/60$$

$$= 10$$

Once found, $\bar{D} = D - \mathbf{1} * \mu$

$$\bar{D} = \begin{bmatrix} 19 & 12 \\ 22 & 6 \\ 6 & 9 \\ 3 & 15 \\ 2 & 13 \\ 20 & 5 \end{bmatrix} - \begin{bmatrix} 12 & 10 \\ 12 & 10 \\ 12 & 10 \\ 12 & 10 \\ 12 & 10 \\ 12 & 10 \end{bmatrix}$$

$$\bar{D} = \begin{bmatrix} 7 & 2 \\ 10 & -4 \\ -6 & -1 \\ -9 & 5 \\ -10 & 3 \\ 8 & -5 \end{bmatrix}$$

2)

Finding the Covariance of a centered matrix is $C_x = \frac{1}{n-1} * \bar{D}^T * \bar{D}$. Where n is the number of objects in the matrix.

$$C_x = \frac{1}{6-1} * \begin{bmatrix} 7 & 10 & -6 & -9 & -10 & 8 \\ 2 & -4 & -1 & 5 & 3 & -5 \end{bmatrix} * \begin{bmatrix} 7 & 2 \\ 10 & -4 \\ -6 & -1 \\ -9 & 5 \\ -10 & 3 \\ 8 & -5 \end{bmatrix}$$

$$C_x = \frac{1}{5} * \begin{bmatrix} 430 & -135 \\ -135 & 80 \end{bmatrix}$$

$$C_x = \begin{bmatrix} 86 & -27 \\ -27 & 16 \end{bmatrix}$$

3) To find the characteristics poolynomials we have to take the determinate of the Covariance minus the eigen values. $\det(C_x - \lambda * 1)$

$$\det\left(\begin{bmatrix} 86 & -27 \\ -27 & 16 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = 0$$

$$\det\left(\begin{bmatrix} 86 - \lambda & -27 \\ -27 & 16 - \lambda \end{bmatrix}\right) = 0$$

$$(86 - \lambda)(16 - \lambda) - (-27)(-27) = 0$$

$$1376 - 102\lambda + \lambda^2 - 729 = 0$$

$$\lambda^2 - 102\lambda + 647 = 0$$

From the polynomial $\lambda^2 - 102\lambda + 647$, the components a, b and c are 1, -102, and 647 respectively. To find the Eigen Values, we find the root of the polynomials. Which can be done with the quadratic equation. Giving us 2 roots; $\lambda_1 = 95.204072$ and $\lambda_2 = 6.795928$.

4)The principal component matrix is made up from the Eigen Vectors of C_x . Using R the Principal Component matrix is

$$\begin{bmatrix} -0.9465153 & -0.3226591 \\ 0.3226591 & -0.9465153 \end{bmatrix}$$

5)The first principal component p_1 is the chosen component divided by the sum of all the components. So the percentage is 0.9333733.

6)The PCA transformation is found by $\mathbf{Y} = \mathbf{P}^{-1} * \mathbf{X}$.

$$\begin{aligned} \mathbf{P}^{-1} &= \begin{bmatrix} -0.9465153 & 0.3226591 \\ -0.3226591 & -0.9465153 \end{bmatrix} \\ \mathbf{Y} &= \begin{bmatrix} 19 & 12 \\ 22 & 6 \\ 6 & 9 \\ 3 & 15 \\ 2 & 13 \\ 20 & 5 \end{bmatrix} * \begin{bmatrix} -0.9465153 & 0.3226591 \\ -0.3226591 & -0.9465153 \end{bmatrix} \\ &= \begin{bmatrix} -14.111881 & -17.48871 \\ -18.887381 & -12.77759 \\ -2.775160 & -10.45459 \\ 2.000340 & -15.16571 \\ 2.301537 & -12.95002 \\ -17.317010 & -11.18576 \end{bmatrix} \end{aligned}$$

7)

i)N here is the number of Attributes in the data set. The percetange of variance state how correlated the variables are

ii)The results of i can be intepret that PCA is not helpful at all. Since we want to reduce the number of dimensions and see the best one. We started with N dimensions and still remain with N dimensions. The case for ii is much more favorable. since only one principal component is needed and show a possible correlation, we can reduce from N dimensions to just 1.