Written by Yarden Nativ, mail: nativyardenac@gmail.com

9^{th} Recitation 24.05.23

Dimensionality reduction

PCA

Principal component analysis (PCA) is a technique that is useful for the compression and classification of data. The purpose is to reduce the dimensionality of a data set (sample) by finding a new set of variables, smaller than the original set of variables, that nonetheless retains most of the sample's information.

By information we mean the variation present in the sample, given by the correlations between the original variables. The new variables, called principal components (PCs), are uncorrelated, and are ordered by the fraction of the total information each retains.

PCA mathematic definitions:

- $T(M \times N)$ is a linear transformation
- $X_i(M \times 1)$ is the corresponding eigenvector of λ_i eigenvalue:
$$T(X_i) = \lambda_i X_i$$
- The eigenvalues are found using the characteristic polynomial $|T - \lambda I| = 0$
- The corresponding eigenvectors are found using $(T - \lambda_i I) \cdot X_i = 0$

PCA algorithm:

1. Given the input $Y_{[N \times M]}$ when N- number of trials, M- number of samples in each trial
2. Subtract the mean from each dimension of $X = Y - E(Y)$. For each sample point/variable dimension we subtract the mean (calculated over trials).
3. Derive eigenvalues and eigenvectors of the Covariance Matrix $X^T \cdot X$. No need to normalize the matrix to number of trials N, although the real covariance matrix does so.
4. Sort the eigenvalues in descending order (from highest to lowest) and write the diagonalizing matrix with the adequate eigenvectors $v_i[M \times 1]$. The eigenvectors are the principal components.
5. To find the projection of trial k (The k-th row of matrix X) on the principal component v_i we will calculate the dot product $k \cdot v_i$, which will be a scalar representing the size of the projected trial k on this principal component.

Written by Yarden Nativ, mail: nativyardenac@gmail.com

Class Exercise:

Given a very simplistic set of recordings with 2 samples for each observation:

	t_1	t_2
trial 1	2	6
trial 2	1.5	5
trial 3	2.5	4

The matrix Y is given by:

$$Y = \begin{pmatrix} 2 & 6 \\ 1.5 & 5 \\ 2.5 & 4 \end{pmatrix}$$

The expectation of the Y is the vector of the averages of the columns:

$$E(Y) = \begin{pmatrix} 2 & 5 \end{pmatrix}$$

And the matrix $X = Y - E(Y)$ is:

$$X = \begin{pmatrix} 0 & 1 \\ -0.5 & 0 \\ 0.5 & -1 \end{pmatrix}$$

The Covariance matrix is given by:

$$Cov = X^T \cdot X = \begin{pmatrix} 0 & 1 \\ -0.5 & 0 \\ 0.5 & -1 \end{pmatrix}^T \cdot \begin{pmatrix} 0 & 1 \\ -0.5 & 0 \\ 0.5 & -1 \end{pmatrix} = \begin{pmatrix} 0 & -0.5 & 0.5 \\ 1 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ -0.5 & 0 \\ 0.5 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 2 \end{pmatrix}$$

Now we want to diagonalize the matrix using the characteristic polynomial:

$$|Cov - \lambda I| = 0$$

$$\begin{vmatrix} 0.5 - \lambda & -0.5 \\ -0.5 & 2 - \lambda \end{vmatrix} = 0$$

$$(0.5 - \lambda)(2 - \lambda) - (-0.5) \cdot 0.5 = 0$$

$$\lambda^2 - 2.5 \cdot \lambda + 0.75 = 0$$

$$\lambda_1 = 0.3486 \quad \lambda_2 = 2.1514$$

λ_2 represents 85% of the variability of the original data (using the calculation $\frac{\lambda_i}{\sum_i \lambda_i}$).

To find the projections of the vectors on those eigenvalues we need to find the eigenvectors using $(Cov - \lambda_i I) \cdot V_i = 0$.

For λ_1, we will assume $V_1 = \binom{a}{b}$:

$$\begin{pmatrix} 0.5 - 0.3486 & -0.5 \\ -0.5 & 2 - 0.3486 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} \xrightarrow{First\ row\ multiplication} 0.1514a - 0.5b = 0$$

$$\rightarrow \quad 0.3028a - b = 0 \rightarrow \; b = 0.3028a$$

$$V_1 = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ 0.3028a \end{pmatrix} \rightarrow V_1 = \begin{pmatrix} 1 \\ 0.3028 \end{pmatrix}$$

We want to use an orthonormal system of eigenvectors. Therefore, we will do a normalization of each vector:

$$V_{1,normalized} = \frac{V_1}{\sqrt{a^2+b^2}} = \begin{pmatrix} 0.9571 \\ 0.2898 \end{pmatrix}$$

Similarly:

$$V_{2,normalized} = \begin{pmatrix} -0.2898 \\ 0.9571 \end{pmatrix}$$

The projection of trial 1, $(0,1)$ on the eigenvector V_1 is $(0,1) \cdot \begin{pmatrix} 0.9571 \\ 0.2898 \end{pmatrix} = 0.2898$

Class Discussion: What are the limitations of PCA?

- Assumptions:
  - Mean and variance (moments 1&2) are sufficient statistics.
  - Large variances are important.
  - The covariance matrix is diagonalizable.
- Limitations:
  - Linear transformation.
- Properties:
  - PCs are orthogonal.

Class Exercise:

2007 exam: The shape of spikes recorded in the Animalis Simplistics may be described by:

$$A(n) = \alpha(n)V_1 + \beta(n)V_2$$

(A, V1 & V2 are vectors of length τ).

The values of α are normally distributed $(\mu = 0, \sigma = 1)$, and $\beta(n) = 2\alpha(n)$.

After performing PCA on the spikes, the number k of non-zero eigenvalues:

$A.\ k = 0$

$B.\ k = 1$

$C.\ k = 2$

$D.\ k > 2$

$E.\ k = n$

Solution:

The maximum possible eigenvalues is 2 because the linear combination is given with two vectors. Nevertheless, given that $\beta(n)$ is a linear combination of $\alpha(n)$ we can understand that there is only 1 eigenvalue, therefore the answer is B.

Computationally:

$$A = \alpha V_1 + \beta V_2 = \alpha(V_1 + 2V_2)$$

We'll define:

$$V_1 + 2V_2 \equiv V_3$$

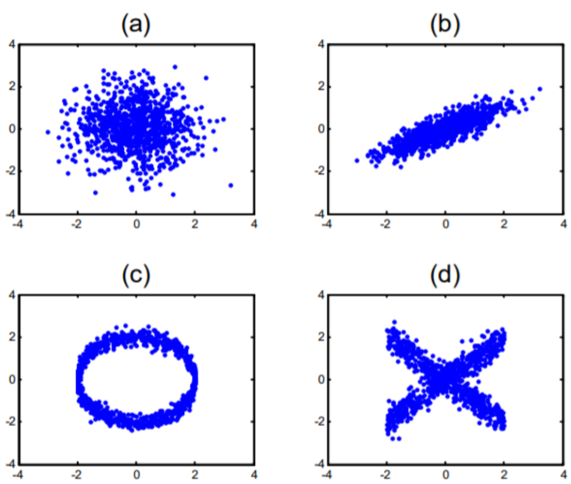
Therefore:

$$A = \alpha V_3$$

All of the information can be represented within one dimension, so there is only one eigenvalue to represent all of the data variability.

Class Exercise:

2006 exam: For which of the following two variable distributions is the first principal component (as calculated using PCA) most useful?



Solution:

The 2-dimensional distribution is B shows the biggest variance along one axis, therefore this data will be best suited for PCA.

Written by Yarden Nativ, mail: nativyardenac@gmail.com

ICA: Independent component analysis

ICA separates the data into statistically independent components. Differently than PCA, the algorithm assumes that the sources of the data are independent and non-Gaussian (at most one can be Gaussian).
Nevertheless, there is no closed solution for ICA algorithms. Commonly, most of the algorithms try to follow one of the possibilities:

- Maximize non-Gaussianity of recovered sources, using:
    - Kurtosis (=0 for a Gaussian)
    - Entropy (maximal for a Gaussian)
- Minimize mutual information between recovered signal components – less information→more independence.
- Maximum likelihood – includes information on priors for the sources.

PCA vs. ICA:

- PCA finds direction of maximal variance
- ICA finds direction of maximal independence in non-Gaussian data (higher-order statistics).
- ICA can be used for blind source separation.

Class Exercise for self-learning:

Exercise: The MEG signal recorded by two Squids (SA & SB) is distributed Normally N(μ=0,σ=1). Given the 3th moment and so on are 0, during three experiments the following covariance matrices were calculated

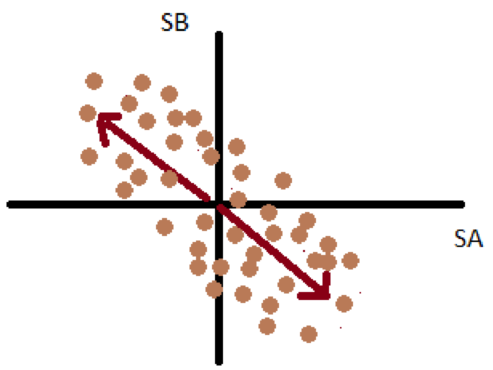
$$Cov_1(SA,SB)=\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix} \qquad Cov_2(SA,SB)=\begin{pmatrix}1 & 0.5\\ 0.5 & 1\end{pmatrix} \qquad Cov_3(SA,SB)=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

a) Sketch sample values of SA vs. SB for the three experiments (demonstrating the joint distribution)
b) Estimate the eigenvalues and eigenvectors for the three experiments
c) In which experiment do the sensors display the highest redundancy and in which one the lowest? Explain.

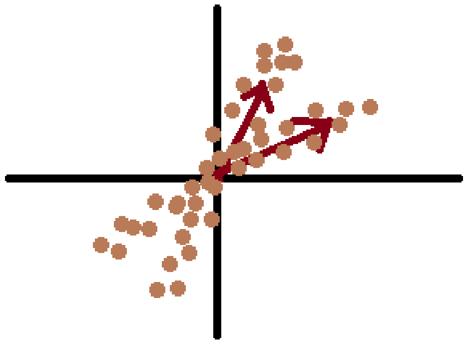
Full Hebrew solution to exercise for self learning:

א. וקטורי העמודות במטריצת $Covariance$ מייצגים את הכיוונים בעלי השונות הגדולה ביותר במרחב בו כל נקודה מייצגת התאמה בין מדידה אחת ב-S_A למדידה אחת באותו זמן ב-S_B. לכן, בהינתן כי שתי הדגימות היו נורמליות, נוכל לצייר את הפיזור על בסיס ציור של וקטורי העמודות על גבי כל ציר.

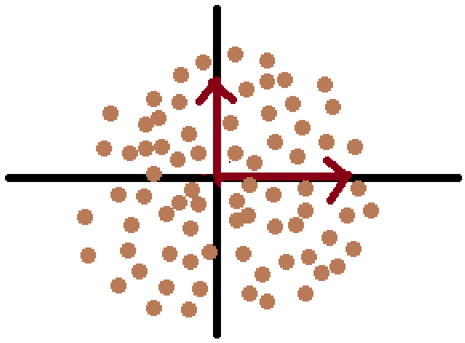
עבור $Cov_1(S_A,S_B)=\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}$ נקבל כי הפיזור המתאים הינו עבור ציר אחד מרכזי:



עבור $Cov_2(S_A,S_B)=\begin{pmatrix}1 & 0.5\\ 0.5 & 1\end{pmatrix}$ נקבל כי הפיזור המתאים הינו עבור שני צירים מרכזיים לא אורתוגונליים זה לזה:



עבור $Cov_3(S_A, S_B) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ נקבל כי הפיזור המתאים הינו עבור שני צירים אורתוגונליים זה לזה, מה שמייצג חוסר תלות:



ב. נחשב את הערכים העצמיים והוקטורים העצמיים עבור כל מטריצה ע״י שימוש בפולינום אופייני.

(1) עבור:

$$Cov_1(S_A, S_B) = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

הפולינום האופייני הינו:

$$P = |Cov_1 - \lambda I| = 0$$

ולכן:

$$\begin{vmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 1 = \lambda^2 - 2\lambda$$

קיבלנו שני ערכים עצמיים:

$\boldsymbol{\lambda_1 = 2 \quad \lambda_2 = 0}$

נציב כל אחד מהם במשוואה $(Cov_1 - \lambda I) \cdot V = 0$, לכן עבור $\lambda_1 = 2$ מתקיים כי הפתרון הוא בדירוג מטריצות:

$$\begin{pmatrix} 1-2 & -1 \\ -1 & 1-2 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} =^{R_2=R_2-R_1} \begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix}$$

ולכן:

$$-V_x - V_y = 0$$

$$V_x = -V_y$$

וזה מתאים לוקטור העצמי $\boldsymbol{V_{\lambda_1=2} = (1, -1)}$.

עבור $\lambda_1 = 0$ מתקיים כי הפתרון הוא בדירוג מטריצות:

$$\begin{pmatrix} 1-0 & -1 \\ -1 & 1-0 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} =^{R_2=R_2+R_1} \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$$

ולכן:

$$V_x - V_y = 0$$

$$V_x = V_y$$

וזה מתאים לוקטור העצמי $\boldsymbol{V_{\lambda_2=0} = (1, 1)}$.

עבור (2):

$$Cov_2(S_A, S_B) = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

הפולינום האופייני הינו:

$$P = |Cov_1 - \lambda I| = 0$$

ולכן:

$$\begin{vmatrix} 1-\lambda & 0.5 \\ 0.5 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 0.25 = \lambda^2 - 2\lambda + 0.75$$

$$\lambda_{1,2} = \frac{2 \pm \sqrt{4 - 4 \cdot 0.75}}{2} = \frac{2 \pm 1}{2}$$

קיבלנו שני ערכים עצמיים:

$$\boldsymbol{\lambda_1 = 1.5 \quad \lambda_2 = 0.5}$$

נציב כל אחד מהם במשוואה $(Cov_1 - \lambda I) \cdot V = 0$, לכן עבור $\lambda_1 = 1.5$ מתקיים כי הפתרון הוא בדירוג מטריצות:

$$\begin{pmatrix} 1-1.5 & 0.5 \\ 0.5 & 1-1.5 \end{pmatrix} = \begin{pmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix} =^{R_2=R_2+R_1} \begin{pmatrix} -0.5 & 0.5 \\ 0 & 0 \end{pmatrix}$$

ולכן:

$$-0.5V_x + 0.5V_y = 0$$

$$V_x = V_y$$

וזה מתאים לוקטור העצמי $\boldsymbol{V_{\lambda_1=1.5} = (1, 1)}$.

עבור $\lambda_1 = 0.5$ מתקיים כי הפתרון הוא בדירוג מטריצות:

$$\begin{pmatrix} 1-0.5 & 0.5 \\ 0.5 & 1-0.5 \end{pmatrix} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} =^{R_2=R_2-R_1} \begin{pmatrix} 0.5 & 0.5 \\ 0 & 0 \end{pmatrix}$$

ולכן:

$$0.5V_x + 0.5V_y = 0$$

$$V_x = -V_y$$

וזה מתאים לוקטור העצמי $\boldsymbol{V_{\lambda_2=0} = (1, -1)}$.

(3) עבור:

$$Cov_3(S_A, S_B) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

המטריצה כבר מלוכסנת ומתקיים עבורה:

$\boldsymbol{\lambda_1 = \lambda_2 = 1}$

והוקטורים העצמיים המתאימים הם וקטורי העמודות, כלומר:

$\boldsymbol{V_{1_{\lambda_1=1}} = (1,0)}$, $\boldsymbol{V_{2_{\lambda_1=1}} = (0,1)}$.

טבלה מסכמת לסעיף זה:

$Cov\ matrix$	$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$\lambda - eigen\ values$	$\lambda_1 = 2$ $\lambda_2 = 0$	$\lambda_1 = 1.5$ $\lambda_2 = 0.5$	$\lambda_1 = \lambda_2 = 1$
$V - eigen\ vectors$	$V_{\lambda_1=2} = (1,-1)$ $V_{\lambda_1=0} = (1,1)$	$V_{\lambda_1=1.5} = (1,1)$ $V_{\lambda_1=0.5} = (1,-1)$	$V_{1_{\lambda_1=1}} = (1,0)$ $V_{2_{\lambda_1=1}} = (0,1)$

ג. באמצעות התשובה לסעיף א׳ וגם התשובה לסעיף ב׳ נוכל לשים לב כי בניסוי הראשון מטריצת ה-Cov_1 היא בעלת שני וקטורים המצביעים על אותו כיוון שונות, וכן במציאת הערכים העצמיים של מטריצה זו התקבל ערך $\lambda = 2$ המצביע על כך שניתן להסביר 100% מהשונות על ידי הוקטור העצמי המתאים $(1,-1)$. לכן בניסוי זה, הניסוי הראשון, הסנסורים מציגים את ה-*redundancy* הגבוהה ביותר.

לעומת זאת בניסוי השלישי, לפנינו מקרה הפוך שבו הוקטורים המצביעים על כיווני השונות מאונכים אחד לשני, ושני הערכים העצמיים של המטריצה הם $\lambda = 1$, מכאן כל אחד מהוקטורים העצמיים $(1,0), (0,1)$ מסביר 50% מהשונות של המידע שנאסף בניסוי. לכן בניסוי השלישי, הסנסורים מציגים את ה-*redundancy* הנמוך ביותר.