This session will be taught in Hybrid format

When joining on Zoom

- When your microphone or camera are switched on you will be able to be heard in the teaching room and may be seen on the screen in the room.
- If the screen is recorded, you may be captured on the recording if you choose to contribute and have your camera switched on.
- Chat messages to everyone and privately to the host may be visible on screen in the room.

When joining in the Teaching room

- Academic teaching staff should let you know when Zoom is connected and you may see the Zoom meeting on the screen.
- Once connected to Zoom, microphones may capture sound from the whole room. Your voice may be shared to Zoom when speaking at a normal talking volume.
- Once connected to Zoom, cameras will capture video images from the room. These may be shared to Zoom. The camera capture will be a wide shot of the whole room or block of seating. It will not capture close-up images of individual students.
- If the session is recorded, video from the room will be captured on the recording. Your voice will be captured on the recording if you choose to contribute.
- If you connect to the Zoom meeting from within the room (for example to use the chat), please keep your audio on mute or use headphones to avoid feedback issues with the room audio.

The recordings may be stored in the cloud ad any personal information within the recording will be processed in accordance to the General Data Protection Regulations 2018.
A substantial contribution is considered to be anything more than merely answering questions or participating in a group discussion. Where you make a substantial contribution to the delivery of the recorded events, a signed consent form will be obtained prior to the recording being made available for viewing. The Consent Form will address your personal informationand any copyright or other intellectual property in the recording.

Dimensional analysis by covariance

Eigendecomposition / PCA

Felipe Orihuela-Espina

14-Oct-2024

Table of Contents

Section 1

Principal Component Analysis (PCA) and Classical Multidimensional Scaling (cMDS)

Principal Component Analysis and Classical Multidimensional Scaling (cMDS)

Let's see PCA and cMDS in action.

Let's start by declaring some data[1];

```
data = [2.5 2.4; ...
0.5 0.7; ...
2.2 2.9; ...
1.9 2.2; ...
3.1 3.0; ...
2.3 2.7; ...
2   1.6; ...
1   1.1; ...
1.5 1.6; ...
1.1 0.9];
```

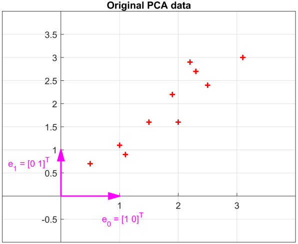

Figure: Original data.

[1]The following slides follow Smith's tutorial [1] with minor adjustments.

Principal Component Analysis and Classical Multidimensional Scaling (cMDS)

Prepare the data for PCA (mean removal);

```
meanadjusted = data -
    mean(data);
```

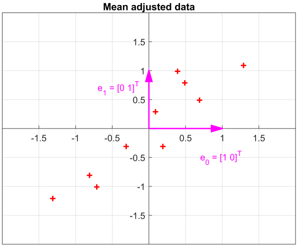

Figure: Mean adjusted data.

Principal Component Analysis and Classical Multidimensional Scaling (cMDS)

Calculate covariances and adjust the eigenvectors from the covariance matrix;

```
covmat = cov(meanadjusted);

eigvalues = eig(covmat);
[normaleigs,D] = eig(covmat);
sorteigvalues =
    sorteigvectors(eigvalues',...
                   eigvalues');
sortnormaleigs =
    sorteigvectors(eigvalues',...
                   normaleigs);
```

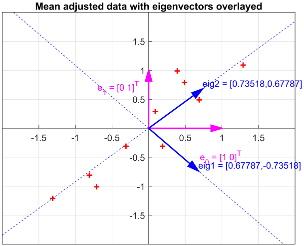

Figure: Eigenvectors

Principal Component Analysis and Classical Multidimensional Scaling (cMDS)

Transform (rotate) the data according to the "new" canonical axes;

```
prefinaldata =
    eigenvectors'*meanadjusted';
    %Final data is the
        weighted sum of
    %the projections
finaldata = prefinaldata';
```

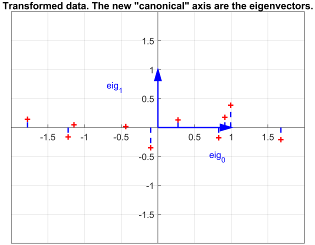

Figure: Transformed (rotated) data.

Principal Component Analysis and Classical Multidimensional Scaling (cMDS)

Show that cMDS is a duality with PCA;

```
%Calculate pairwise distances
    on data
D =
    squareform(pdist(meanadjusted,'
%Feed the pairwise distance
    to cMDS
embeddedPoints = cmdscale(D);
%We may need to flip to get
    the SAME solution that PCA
embeddedPoints(:,1) =
    -1*embeddedPoints(:,1);
```

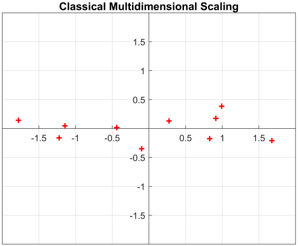

Figure: Classical Multidimensional Scaling (cMDS)

Principal Component Analysis and Classical Multidimensional Scaling (cMDS)

Rule of thumb:

- Higher covariance between variables suggests they carry redundant information, indicating fewer effective dimensions.
- Conversely, low covariance often implies unique contributions, hinting at higher dimensionality.

. . . but beware! This is only a rule of thumb. It is not always the case! Guessing the dimensionality of a dataset directly looking at the covariance matrix is very difficult to interpret, ergo PCA!

Principal Component Analysis (PCA)

Exercise

Let be the multivariate distribution below. Using the estimator with the Bessel's correction, calculate the mean and covariance of the joint distribution.

X_1	X_2	X_3	X_4
6	9	4	4
8	9	5	-4
-7	6	5	9
8	-7	-2	-9
3	-2	3	-1
-8	8	-7	-2
-4	6	4	5
1	9	-9	2
9	3	-4	-6
9	-9	-9	0
-7	6	3.5	-1
9	9	6	3

Exercise

The mean of multivariate distributions is the vector of marginal means, but the (co-)variance has now both the marginal terms and the interaction terms;

$$\mu = \begin{pmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \mu_{X_3} \\ \mu_{X_4} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma^2_{X_1} & cov(X_1, X_2) & cov(X_1, X_3) & cov(X_1, X_4) \\ cov(X_2, X_1) & \sigma^2_{X_2} & cov(X_2, X_3) & cov(X_2, X_4) \\ cov(X_3, X_1) & cov(X_3, X_2) & \sigma^2_{X_3} & cov(X_3, X_4) \\ cov(X_4, X_1) & cov(X_4, X_2) & cov(X_4, X_3) & \sigma^2_{X_4} \end{pmatrix} \tag{1}$$

Principal Component Analysis (PCA)

Exercise

```
>> X =[ ...
    6, 9, 4, 4; ...
    8, 9, 5, -4; ...
   -7,  6, 5, 9; ...
    8, -7, -2, -9;
         ...
    3, -2, 3, -1;
         ...
   -8, 8, -7, -2;
        ...
   -4, 6,  4, 5; ...
    1, 9, -9, 2; ...
    9, 3, -4, -6;
         ...
    9, -9, -9, 0;
         ...
   -7, 6,  3.5, -1;
        ...
    9, 9, 6, 3]
>> mean(X)'
>> cov(X)
```

$$\mu = \begin{pmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \mu_{X_3} \\ \mu_{X_4} \end{pmatrix} \approx \begin{pmatrix} \bar{x_1} \\ \bar{x_2} \\ \bar{x_3} \\ \bar{x_4} \end{pmatrix} = \begin{pmatrix} 2.2500 \\ 3.9167 \\ -0.0417 \\ 0 \end{pmatrix} \tag{2}$$

$$\Sigma = \begin{pmatrix} \sigma^2_{X_1} & cov(X_1, X_2) & cov(X_1, X_3) & cov(X_1, X_4) \\ cov(X_2, X_1) & \sigma^2_{X_2} & cov(X_2, X_3) & cov(X_2, X_4) \\ cov(X_3, X_1) & cov(X_3, X_2) & \sigma^2_{X_3} & cov(X_3, X_4) \\ cov(X_4, X_1) & cov(X_4, X_2) & cov(X_4, X_3) & \sigma^2_{X_4} \end{pmatrix} \tag{3}$$

$$\approx \begin{pmatrix} s^2_{X_1} & s_{X_1,X_2} & s_{X_1,X_3} & s_{X_1,X_4} \\ s_{X_2,X_1} & s^2_{X_2} & s_{X_2,X_3} & s_{X_2,X_4} \\ s_{X_3,X_1} & s_{X_3,X_2} & s^2_{X_3} & s_{X_3,X_4} \\ s_{X_4,X_1} & s_{X_4,X_2} & s_{X_4,X_3} & s^2_{X_4} \end{pmatrix} \tag{4}$$

$$= \begin{pmatrix} 48.5682 & -16.1591 & -3.0341 & -15.2727 \\ -16.1591 & 41.3561 & 13.8144 & 14.0000 \\ -3.0341 & 13.8144 & 33.6572 & 10.0455 \\ -15.2727 & 14.0000 & 10.0455 & 24.9091 \end{pmatrix} \tag{5}$$

Thanks

Thank you! Questions?

References I

Lindsay I. Smith.
A tutorial on principal component analysis, FEB 2002.