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.

Consolidation week

Exercises

Felipe Orihuela-Espina

8-Nov-2024

Table of Contents

Section 1

Dimensional analysis by statistical independence: ICA

ICA

Example

Let be the dataset $X \in \mathbb{R}^2$ expressed in the canonical basis, and the coordinate basis A and B;

$$X^T = \begin{bmatrix} 0.7200 & 3.1200 \\ -5.4800 & -3.6800 \\ 2.6200 & -1.1800 \\ 3.4200 & -3.4800 \\ 0.9200 & -2.9800 \\ 1.7200 & 4.2200 \\ 1.5200 & 2.9200 \\ -1.9800 & -0.7800 \\ 0.7200 & 5.5200 \\ -4.1800 & -3.6800 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 4 \\ -2 & 3 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 7 \\ -2 & 4 \end{bmatrix}$$

Watch out! This is the transpose of X.

Which basis is more relevant according to ICA?

ICA

Example

Solution:

ICA is based on the principle of (statistical) independence among the sources i.e. the coordinate basis. So the exercise is fundamentally asking that we compute the statistical independence for each basis and then picking the basis that is more independent.

In principle, because we are not being asked to simulate the iterative optimization, we do not need to build a cost function from the entropy of the cumulative joint distribution, and we can work directly with the joint distribution.

Example

Solution (Cont.):

We know that two variables are (statistically) independent when the joint probability is equal to the product of the marginal probabilities. That is:

$$Pr(X_1, X_2) = Pr(X_1)Pr(X_2)$$

with X_1 and X_2 the columns of X^T. 0.3cm

Or in other words, when knowing about X_1 does not give us information about X_2.

Watch out!

The joint probability distribution $f_{X_1,X_2} = Pr(X_1, X_2)$ is not a single scalar! Neither are the marginals $f_{X_1} = Pr(X_1)$ and $f_{X_2} = Pr(X_2)$. In their discretized approximations, $f_{X_1,X_2} = Pr(X_1, X_2)$ is a matrix, and the marginals $f_{X_1} = Pr(X_1)$ and $f_{X_2} = Pr(X_2)$ are vectors.

Example

Solution (Cont.):
We can therefore estimate how close the variables X_1 and X_2 are to being independent by building an error function:

$$e(X_1, X_2) = \|Pr(X_1, X_2) - Pr(X_1)Pr(X_2)\|_2$$

The closer $e(X_1, X_2)$ is to 0, the closer X_1 and X_2 are to being independent.

Example

Solution (Cont.):

So here is the rationale for our solution;

1. Re-express the X canonical coordinates into A or B, X_A or X_B respectively.
    - This will require the computation of the inverse matrices A^{-1} and B^{-1}.
2. Calculate $Pr(X_{A,1})$, $Pr(X_{A,2})$ and their joint $Pr(X_{A,1}, X_{A,2})$, and analogously $Pr(X_{B,1})$, $Pr(X_{B,2})$ and their joint $Pr(X_{B,1}, X_{B,2})$
    - As X is dense (e.g. $X \in \mathbb{R}^2$), we shall need to define the interval events in order to calculate probabilities.
3. Calculate the errors $e_A = \|Pr(X_{A,1}, X_{A,2}) - Pr(X_{A,1})Pr(X_{A,2})\|_2$ and $e_B = \|Pr(X_{B,1}, X_{B,2}) - Pr(X_{B,1})Pr(X_{B,2})\|_2$.
4. Choose A or B depending on $min(e_A, e_B)$.

Example

Solution (Cont.):

1. **Re-express the X canonical coordinates into A or B, X_A or X_B respectively.**

Before we can calculate the probabilities for our distributions, we need to re-express X in both A and B.

To transform X canonical coordinates into A or B, X_A or X_B, respectively, we solve the following system of equations;

$$\begin{aligned} IX = AX_A \quad &\Rightarrow \quad A^{-1}X = X_A \\ IX = BX_B \quad &\Rightarrow \quad B^{-1}X = X_B \end{aligned}$$

Example

Solution (Cont.):

Therefore, we start by calculating the inverses A^{-1} or B^{-1};

$$A^{-1} = \frac{adj(A)}{det(A)} = \frac{cof(A)^T}{det(A)} = \begin{bmatrix} 0.2727 & -0.3636 \\ 0.1818 & 0.0909 \end{bmatrix}$$

$$B^{-1} = \frac{adj(B)}{det(B)} = \frac{cof(B)^T}{det(B)} = \begin{bmatrix} 0.2857 & -0.5000 \\ 0.1429 & 0 \end{bmatrix}$$

Example

Solution (Cont.):
And with the inverses A^{-1} or B^{-1}, we can now calculate the new coordinates X_A or X_B;

Rounding to 2 significant figures;

$$X_A = A^{-1}X = \begin{bmatrix} -0.94 & -0.16 & 1.14 & 2.20 & 1.33 & -1.07 & -0.65 & -0.26 & -1.81 & 0.20 \\ 0.41 & -1.33 & 0.37 & 0.31 & -0.10 & 0.70 & 0.54 & -0.43 & 0.63 & -1.09 \end{bmatrix}$$

$$X_B = B^{-1}X = \begin{bmatrix} -1.35 & 0.27 & 1.34 & 2.72 & 1.75 & -1.62 & -1.03 & -0.18 & -2.55 & 0.6500 \\ 0.10 & -0.78 & 0.37 & 0.49 & 0.13 & 0.25 & 0.22 & -0.28 & 0.10 & -0.6000 \end{bmatrix}$$

Example

Solution (Cont.):

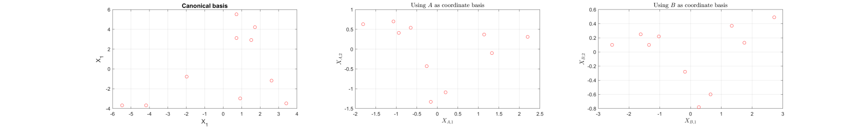


Figure: Scatter plot of the dataset X under the three coordinate basis; left) canonical, middle) A and right) B.

Example

Solution (Cont.):

2. **Calculate** $Pr(X_{A,1})$, $Pr(X_{A,2})$ **and their joint** $Pr(X_{A,1}, X_{A,2})$

As aforementioned, since X is dense (e.g. $X \in \mathbb{R}^2$), we need to define the interval events in order to calculate probabilities.

There are several ways to compute this step, but the easiest is to use equispaced binning, i.e. a number of intervals of the same size.

- For no particular reason, I will use 5 bins in this solution for all $X_{A,1}$, $X_{A,2}$, $X_{B,1}$ and $X_{B,2}$, but feel free to use a different number of intervals, or even another approach!

Example

Solution (Cont.):

Here are the intervals;

	min	max	range	bin size
$X_{A,1}$	-1.81	2.20	4.01	0.80
$X_{A,2}$	-1.33	0.70	2.03	0.41
$X_{B,1}$	-2.55	2.72	5.27	1.05
$X_{B,2}$	-0.78	0.49	1.27	0.25

bin	$X_{A,1}$	$X_{A,2}$	$X_{B,1}$	$X_{B,2}$
1	[-1.81,-1.01)	[-1.33,-0.92)	[-2.55,-1.5)	[-0.78,-0.53)
2	[-1.01,-0.21)	[-0.92,-0.51)	[-1.5,-0.45)	[-0.53,-0.28)
3	[-0.21,0.59)	[-0.51,-0.1)	[-0.45,0.6)	[-0.28,-0.03)
4	[0.59,1.39)	[-0.1,0.31)	[0.6,1.65)	[-0.03,0.22)
5	[1.39,2.19]	[0.31,0.72]	[1.65,2.7]	[0.22,0.47]

Example

Solution (Cont.):

Note:

- The last intervals are closed on both ends.
- The edges of the intervals have a rounding error; Given the small number of points in our point cloud it is convenient to alleviate this by replacing by the exact maxima of the variable at least in the last bin;

bin	$X_{A,1}$	$X_{A,2}$	$X_{B,1}$	$X_{B,2}$
int_1	[-1.81,-1.01)	[-1.33,-0.92)	[-2.55,-1.5)	[-0.78,-0.53)
int_2	[-1.01,-0.21)	[-0.92,-0.51)	[-1.5,-0.45)	[-0.53,-0.28)
int_3	[-0.21,0.59)	[-0.51,-0.1)	[-0.45,0.6)	[-0.28,-0.03)
int_4	[0.59,1.39)	[-0.1,0.31)	[0.6,1.65)	[-0.03,0.22)
int_5	[1.39,2.20]	[0.31,0.70]	[1.65,2.72]	[0.22,0.49]

These intervals represent our events for calculating the probabilities.

Example

Solution (Cont.):
For the marginal probabilities $Pr(X_{A,1})$, $Pr(X_{A,2})$, $Pr(X_{B,1})$ and $Pr(X_{B,2})$, this is merely a counting exercise and a division by the number of points in the cloud ($n = 10$);

$Pr(X \in int_j)$	$X_{A,1}$	$X_{A,2}$	$X_{B,1}$	$X_{B,2}$
$Pr(X \in int_1)$	0.2	0.2	0.2	0.2
$Pr(X \in int_2)$	0.3	0	0.2	0.1
$Pr(X \in int_3)$	0.2	0.1	0.2	0
$Pr(X \in int_4)$	0.2	0.1	0.2	0.4
$Pr(X \in int_5)$	0.1	0.6	0.2	0.3

Example

Solution (Cont.):

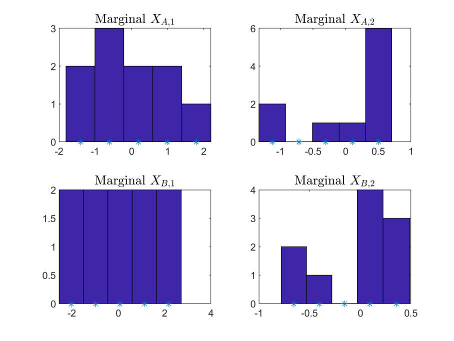


Figure: Marginal distributions for $X_{A,1}$, $X_{A,2}$, $X_{B,1}$, $X_{B,2}$ respectively. Note how the marginal distribution in $P(X_{B,1})$ is a uniform distribution.

Example

Solution (Cont.):
For the joint probabilities however, we need to *cross* (Cartesian product) all pairs of events.

For X_A, we *cross* the events from $X_{A,1}$ with those in $X_{A,2}$

$Pr(X_{A,1}, X_{A,2})$	$Pr(X_{A,2} \in int_1)$	$Pr(X_{A,2} \in int_2)$	$Pr(X_{A,2} \in int_3)$	$Pr(X_{A,2} \in int_4)$	$Pr(X_{A,2} \in int_5)$
$Pr(X_{A,1} \in int_1)$	0	0	0	0	0.2
$Pr(X_{A,1} \in int_2)$	0	0	0.1	0	0.2
$Pr(X_{A,1} \in int_3)$	0.2	0	0	0	0
$Pr(X_{A,1} \in int_4)$	0	0	0	0.1	0.1
$Pr(X_{A,1} \in int_5)$	0	0	0	0	0.1

Example

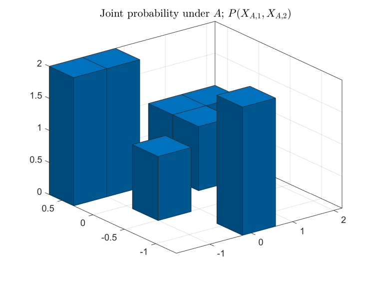
Solution (Cont.):

And for X_B, we *cross* the events from $X_{B,1}$ with those in $X_{B,2}$

$Pr(X_{B,1}, X_{B,2})$	$Pr(X_{B,2} \in int_1)$	$Pr(X_{B,2} \in int_2)$	$Pr(X_{B,2} \in int_3)$	$Pr(X_{B,2} \in int_4)$	$Pr(X_{B,2} \in int_5)$
$Pr(X_{B,1} \in int_1)$	0	0	0	0.1	0.1
$Pr(X_{B,1} \in int_2)$	0	0	0	0.1	0.1
$Pr(X_{B,1} \in int_3)$	0.1	0	0.1	0	0
$Pr(X_{B,1} \in int_4)$	0.1	0	0	0	0.1
$Pr(X_{B,1} \in int_5)$	0	0	0	0.1	0.1

Example

Solution (Cont.):



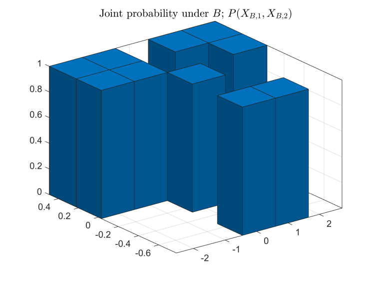


Figure: Joint distributions under (left) coordinate basis A, $P(X_{A,1}, X_{A,2})$, and (right) coordinate basis B, $P(X_{B,1}, X_{B,2})$. Note how the pattern in $P(X_{B,1}, X_{B,2})$ resembles more a uniform distribution than the pattern in $P(X_{A,1}, X_{A,2})$.

ICA

Example

Solution (Cont.):

3. **Calculate the errors** $e_A = \|Pr(X_{A,1}, X_{A,2}) - Pr(X_{A,1})Pr(X_{A,2})\|_2$ **and** $e_B = \|Pr(X_{B,1}, X_{B,2}) - Pr(X_{B,1})Pr(X_{B,2})\|_2$

In order to calculate the errors, we already have the joint probability distributions $Pr(X_{A,1}, X_{A,2})$ and $Pr(X_{B,1}, X_{B,2})$, but we need to calculate the products of the marginals first;

$$Pr(X_{A,2})Pr(X_{A,1})^T = \begin{bmatrix} 0.04 & 0.06 & 0.04 & 0.04 & 0.02 \\ 0 & 0 & 0 & 0 & 0 \\ 0.02 & 0.03 & 0.02 & 0.02 & 0.01 \\ 0.02 & 0.03 & 0.02 & 0.02 & 0.01 \\ 0.12 & 0.18 & 0.12 & 0.12 & 0.06 \end{bmatrix}$$

$$Pr(X_{B,2})Pr(X_{B,1})^T = \begin{bmatrix} 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \\ 0.02 & 0.02 & 0.02 & 0.02 & 0.02 \\ 0 & 0 & 0 & 0 & 0 \\ 0.08 & 0.08 & 0.08 & 0.08 & 0.08 \\ 0.06 & 0.06 & 0.06 & 0.06 & 0.06 \end{bmatrix}$$

Note the pattern in $Pr(X_{B,1})Pr(X_{B,2})^T$ (all columns are equal), you can almost guess that the error under B is going to be smaller!

Example

Solution (Cont.):
We can now calculate the errors;

$$
\begin{aligned}
e_A = e(X_{A,1}, X_{A,2}) &= \left\| Pr(X_{A,1}, X_{A,2}) - Pr(X_{A,1})Pr(X_{A,2}) \right\|_2 \\
&= \sqrt{\sum_i \sum_j \left(Pr(X^i_{A,1}, X^j_{A,2}) - Pr(X^i_{A,1})Pr(X^j_{A,2}) \right)^2} \\
&= 0.6
\end{aligned}
$$

$$
\begin{aligned}
e_B = e(X_{B,1}, X_{B,2}) &= \left\| Pr(X_{B,1}, X_{B,2}) - Pr(X_{B,1})Pr(X_{B,2}) \right\|_2 \\
&= \sqrt{\sum_i \sum_j \left(Pr(X^i_{B,1}, X^j_{B,2}) - Pr(X^i_{B,1})Pr(X^j_{B,2}) \right)^2} \\
&= 0.3162
\end{aligned}
$$

Example

Solution (Cont.):

4. **Choose A or B depending on $min(e_A, e_B)$.**

Finally, we only need to check which one of the errors is smaller.

$$e_A = 0.6$$
$$e_B = 0.3162$$

In this case, this is the one under coordinate basis B.

In other words, the basis that form the coordinate basis B are (virtually) independent from each other when expressing the dataset X.

ICA

Example

Solution (Cont.):

If you want to check in MATLAB, see the companion code file

`ConsolidationWeek_Tutorial_ICAExercise.m`

Thanks

Thank you! Questions?