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

4-Nov-2024

Table of Contents

Section 1

Nature of the Data: Representation

Example

Using the 7-bit ASCII code as numerical representation, tokenizing words, and using the orthonormal projection based in the normalized Dirac's delta $\mathbf{p}[n] = \sum_i p[i]\delta[n-i]$ with $\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$, represent the following message as a point cloud $P = [\mathbf{p_1} \ldots \mathbf{p_k}]$ and present its scatterplot:

"The cat ate hot pie"

Dec	Char	Dec	Char	Dec	Char	Dec	Char
0	Ctrl-@ NUL	32	Space	64	@	96	`
1	Ctrl-A SOH	33	!	65	A	97	a
2	Ctrl-B STX	34	"	66	B	98	b
3	Ctrl-C ETX	35	#	67	C	99	c
4	Ctrl-D EOT	36	$	68	D	100	d
5	Ctrl-E ENQ	37	%	69	E	101	e
6	Ctrl-F ACK	38	&	70	F	102	f
7	Ctrl-G BEL	39	'	71	G	103	g
8	Ctrl-H BS	40	(	72	H	104	h
9	Ctrl-I HT	41	)	73	I	105	i
10	Ctrl-J LF	42	*	74	J	106	j
11	Ctrl-K VT	43	+	75	K	107	k
12	Ctrl-L FF	44	,	76	L	108	l
13	Ctrl-M CR	45	-	77	M	109	m
14	Ctrl-N SO	46	.	78	N	110	n
15	Ctrl-O SI	47	/	79	O	111	o
16	Ctrl-P DLE	48	0	80	P	112	p
17	Ctrl-Q DCI	49	1	81	Q	113	q
18	Ctrl-R DC2	50	2	82	R	114	r
19	Ctrl-S DC3	51	3	83	S	115	s
20	Ctrl-T DC4	52	4	84	T	116	t
21	Ctrl-U NAK	53	5	85	U	117	u
22	Ctrl-V SYN	54	6	86	V	118	v
23	Ctrl-W ETB	55	7	87	W	119	w
24	Ctrl-X CAN	56	8	88	X	120	x
25	Ctrl-Y EM	57	9	89	Y	121	y
26	Ctrl-Z SUB	58	:	90	Z	122	z
27	Ctrl-[ESC	59	;	91	[	123	{
28	Ctrl- FS	60	<	92	₩	124	¦
29	Ctrl-] GS	61	=	93	]	125	}
30	Ctrl-^ RS	62	>	94	^	126	~
31	Ctrl_ US	63	?	95	_	127	DEL

Figure: 7-bits ASCII code.

Representation

Example

Solution (Cont.):
Let's start by switching the text to its numerical representation using the ASCII code. This only involves switching each character by its ASCII decimal representation:

[84, 104, 101, 32, 99, 97, 116, 32, 97, 116, 101, 32, 104, 111, 116, 32, 112, 105, 101]

Example

Solution (Cont.):
The next step it tokenizing by words. Not that this involves splitting words and removing the white spaces (32).

$$\begin{matrix} 84 & 99 & 97 & 104 & 112 \\ 104 & 97 & 116 & 111 & 105 \\ 101 & 116 & 101 & 116 & 101 \end{matrix}$$

We represent as columns because vectors $\mathbf{p_k}$ in P are indicated as columns. Therefore, since there are 5 words, I have 5 columns and since each word has 3 letters I have 3 rows.

Example

Solution (Cont.):
Strictly speaking, the afore matrix are just elements of a set, but they are not yet expressed in a vector space. For all that we know, these numerical figures may as well be pictorial representations. They do not carry any meaning for operational purposes.

Hence, we continue by *projecting* these elements into a (vector) space. We are told to use the orthonormal projection $\mathbf{p}[n] = \sum_i p[i]\delta[n-i]$ with the normalized Dirac's delta;

$$\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

Example

Solution (Cont.):
This specific projection is "transparent" from the numerical representation point of view; since each coordinate gets its same representation and the order of the dimensions is maintained.

For instance, is we want to transform the second word $w_2 = \begin{bmatrix} 99 \\ 97 \\ 116 \end{bmatrix}$ to a point in a space $\mathbf{p}_2$; we get

$$\begin{aligned} \mathbf{p}_2[n] =& w_2[1]\delta[n-1] + w_2[2]\delta[n-2] + w_2[3]\delta[n-3] \\ =& 99 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 97 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 116 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{aligned}$$

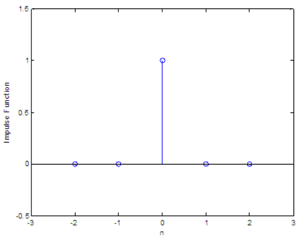
Hence;

$$\mathbf{p}_2 = \begin{bmatrix} 99 \\ 97 \\ 116 \end{bmatrix}$$

Representation

Example

Solution (Cont.):
Although not needed to solve the exercise if you are confused by the $n-k$ have a look's at Dirac's comb. Imagine "pulling" the impulse at location k to the origin. That is if you want to bring the 3rd impulse, $k=3$ to need to shift the $\delta[n]$ function by $-3=-k$, i.e. $\delta[n-k]$.



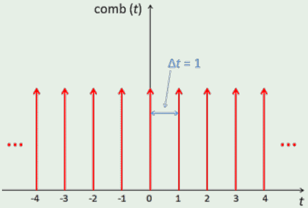


Figure: Dirac's delta (left) and comb (right). Right figure reproduced from [Seeber and Ulrici (2016) ChemTexts 2:18]

In other words, in every instance of $\delta[n-k]$ you "only" sample the k-th element, because the δ function will be 0 everywhere else.

Representation

Example

Solution (Cont.):

Ergo, by projecting the ASCII matrix to a space using the orthonormal projection based in Dirac's delta we get;

$$P = \begin{pmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 & \mathbf{p}_5 \end{pmatrix} = \begin{pmatrix} 84 & 99 & 97 & 104 & 112 \\ 104 & 97 & 116 & 111 & 105 \\ 101 & 116 & 101 & 116 & 101 \end{pmatrix}$$

...which is of course the same numerical representation, but now those numbers have a proper numerical meaning.

Remark

In most sources, this step is considered trivial (well...it is in the mathematical sense) and hence the step is entirely skip. I'm only including explictly in this exercise for completeness because the orthonormal projection based in Dirac's delta is not the only projection that can be used, and for other projections, the numerical representation may actually change.

Example

Solution (Cont.):

The final step is the representation of the scatter plot of this point cloud:

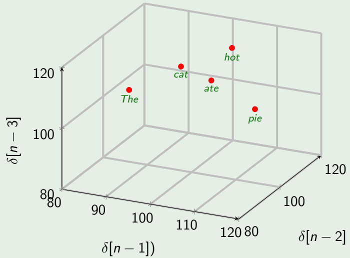


Figure: Representation of "The cat ate hot pie" as a point cloud in a 3D Euclidean space using the orthonormal projection based in Dirac's delta over the ASCII decimal representation.

Section 2

Nature of the Data: Change of coordinate basis

Change of coordinates basis

Example

Let be two coordinate basis A and B

$$A = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \quad B = \left\{ \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right\}$$

And let be the point $\mathbf{p} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ expressed in the coordinate basis A.

Find out the coordinates of $\mathbf{p}$ in the coordinate basis B.

Change of coordinates basis

Example

Solution:

If it helps, we can start by plotting both, the basis as well as the point.

In the canonical coordinates, the point $\mathbf{p}$ will be[a];

$$\mathbf{p} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \cdot 1 + 0 \cdot -2 \\ 1 \cdot 1 + 1 \cdot -2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

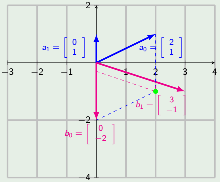


Figure: Change of coordinate basis.

[a] Note that I'm not explicitly pre-multiplying by the inverse of the canonical basis. This is becasue the matrix representing the canonical basis is the identity matrix, for which its inverse is also the identity matrix.

Change of coordinates basis

Example

Solution (Cont.):

The solution consists of establishing a system of equations;

$$\mathbf{p} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} p_{b1} \\ p_{b2} \end{bmatrix}$$

or in matrix form:

$$\mathbf{p} = Ap_A = Bp_B$$

From this later, we can clear p_B;

$$Bp_B = Ap_A \quad \Rightarrow \quad B^{-1}Bp_B = B^{-1}Ap_A \quad \Rightarrow \quad Ip_B = B^{-1}Ap_A \quad \Rightarrow \quad p_B = B^{-1}Ap_A$$

We therefore need to calculate B^{-1} first.

Change of coordinates basis

Example

Solution (Cont.):

To calculate B^{-1} we can use the adjoint method.

1. Calculate $det(B)$: $det(B) = 0 \cdot -1 - 3 \cdot -2 = 6$
2. Calculate $cof(B)$: $cof(B) = \begin{bmatrix} -1 & -2 \\ 3 & 0 \end{bmatrix}$
3. Calculate $adj(B)$: $adj(B) = cof(B)^T = \begin{bmatrix} -1 & 3 \\ -2 & 0 \end{bmatrix}$
4. Calculate B^{-1}: $B^{-1} = \frac{adj(B)}{det(B)} = \begin{bmatrix} -1/6 & 3/6 \\ -2/6 & 0/6 \end{bmatrix} = \begin{bmatrix} -1/6 & 0.5 \\ -0.33 & 0 \end{bmatrix}$

Change of coordinates basis

Example

Solution (Cont.):
Finally, we can solve;

$$p_B = B^{-1} A p_A = \begin{bmatrix} -1/6 & 0.5 \\ -0.33 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 0.16 \\ 0.66 \end{bmatrix}$$

Section 3

Dimensional analysis by covariance; PCA

Eigendecomposition

Example

For $\mathbb{R}^2$, plot the value of the eigenvalues as a function of the correlation of X_1 and X_2, $corr(X_1, X_2)$

Eigendecomposition

Example

For $\mathbb{R}^2$, plot the value of the eigenvalues as a function of the correlation of X_1 and X_2, $corr(X_1, X_2)$

Solution:
The exercise is perhaps a bit long to compute by hand but otherwise conceptually simple as the solution only involves the calculation of the eigendecomposition for several values of the correlation of X_1 and X_2, $corr(X_1, X_2)$.

Eigendecomposition

Example

Solution (Cont.):
Interestingly, to solve the exercise we do not even need to calculate the covariances or the correlations. All we need to start is to remember the structure of the covariance matrix;

$$Cov(X_1, X_2) = \begin{pmatrix} Var(X_1) & cov(X_1, X_2) \\ cov(X_2, X_1) & Var(X_2) \end{pmatrix}$$

and we know that the correlation is just the standardization of the covariance;

$$Corr(X_1, X_2) = \begin{pmatrix} 1 & corr(X_1, X_2) \\ corr(X_2, X_1) & 1 \end{pmatrix}$$

Example

Solution (Cont.):

Note that the correlation matrix:

- It is ALSO a covariance matrix. It is the covariance matrix that you will get if X_1 and X_2 had zero mean and standard deviation 1. Ergo, we can safely feed the correlation matrix directly to the eigendecomposition for PCA.
- Its main diagonal is always 1.
- Its symmetric, i.e. $corr(X_1, X_2) = corr(X_2, X_1)$.

Therefore, in the 2×2 case, the whole correlation matrix (and hence the covariance matrix) only has one parameter; either $corr(X_1, X_2)$ or $corr(X_2, X_1)$.

Eigendecomposition

Example

Solution (Cont.):

We further know that the correlation (standardized covariance) goes from -1 to 1.

So we are looking to create a plot in which;

- the abcisa axis goes from -1 to 1 representing $corr(X_1, X_2)$, and
- the ordinate axis represents 2 functions corresponding to both eigenvalues.

Example

Solution (Cont.):

The exercise itself, requires;

1. Pick a few values of $corr(X_1, X_2) \in [-1, 1]$.
2. For each selected value of $corr(X_1, X_2) = a$ build the correlation matrix; $Corr(X_1, X_2) = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}$
3. Solve the eigendecomposition for each the correlation matrices
4. Plot the values.

Eigendecomposition

Example

Solution (Cont.):

1. Pick a few values of $corr(X_1, X_2) \in [-1, 1]$.
   - For completeness you want to include at least the boundaries of the interval themselves, i.e. -1 and 1, as well as the case $corr(X_1, X_2) = 0$ as we know that in correlation as well as covariance the sign indicates the direction of the relation but the magnitude is what encodes the strength of the relation and hence 0 represents the case where this relation is minimal

So for instance we can choose;

$$corr(X_1, X_2) = \{-1, -0.5, 0, 0.5, 1\}$$

If solving with a computer, feel free to pick a few more points for a smoother curve.

Eigendecomposition

Example

Solution (Cont.):

2. For each selected value of $corr(X_1, X_2) = a$ build the correlation matrix.

$$Corr(X_1, X_2)|_{a=-1} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad Corr(X_1, X_2)|_{a=-0.5} = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix} \quad Corr(X_1, X_2)|_{a=0} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$Corr(X_1, X_2)|_{a=0.5} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \quad Corr(X_1, X_2)|_{a=1} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Example

Solution (Cont.):

3. Solve the eigendecomposition for each the correlation matrices
   - Since we only need the eigenvalues, we do not even need to calculate the eigenvectors! Saving you a good amount of computation.
   - We can get the eigenvalues solving $det(A - \lambda I) = 0$ leading to the characteristic polynomial of A.

Eigendecomposition

Example

Solution (Cont.):

3. Solve the eigendecomposition for each the correlation matrices

For $corr(X_1, X_2) = -1$

$$(A - \lambda I) = \begin{pmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{pmatrix}$$

Hence;

$$det(A - \lambda I) = (1-\lambda)(1-\lambda) - (-1)(-1) = 1 - 2\lambda + \lambda^2 - 1 = \lambda^2 - 2\lambda$$

... and equating this to 0;

$$\lambda^2 - 2\lambda = 0 \quad \Rightarrow \quad \lambda_1 = 2,\ \lambda_2 = 0$$

Now, proceed the same for the other chosen values of $corr(X_1, X_2)$

Eigendecomposition

Example

Solution (Cont.):

3. Solve the eigendecomposition for each the correlation matrices

For $corr(X_1, X_2) = -0.5$

$$(A - \lambda I) = \begin{pmatrix} 1-\lambda & -0.5 \\ -0.5 & 1-\lambda \end{pmatrix}$$

Hence;

$$det(A - \lambda I) = (1-\lambda)(1-\lambda) - (-0.5)(-0.5) = 1 - 2\lambda + \lambda^2 - 0.25 = \lambda^2 - 2\lambda + 0.75$$

. . . and equating this to 0;

$$\lambda^2 - 2\lambda + 0.75 = 0 \quad \Rightarrow \quad \lambda_1 = 1.5,\ \lambda_2 = 0.5$$

Can you do the rest on your own?

Eigendecomposition

Example

Solution (Cont.):

4. Finally, plot the values

a	λ_1	λ_0
-1	2	0
-0.5	1.5	0.5
0	1	1
0.5	1.5	0.5
1	2	0

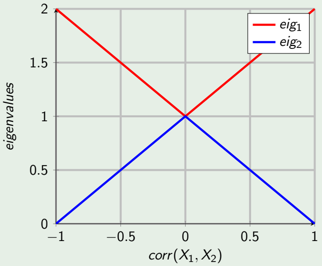


Figure: Behaviour of the eigenvalues as a function of the $corr(X_1, X_2)$ in $\mathbb{R}^2$.

Eigendecomposition

Example

Solution (Cont.):

And if you want to quick check in MATLAB these eigen values:

```
eig([1 -1; -1 1])
eig([1 -0.5; -0.5 1])
eig([1 0; 0 1])
eig([1 0.5; 0.5 1])
eig([1 1; 1 1])
```

Section 4

Dimensional analysis by (co-)occurrence: Latent Semantic Analysis

Latent Semantic Analysis

Example

Let be an occurrence matrix A representing *words* $\times$ *documents*, and its corresponding outcome of the Latent Semantic Analysis over its normalized (zero-mean and standard deviation 1) version A_z:

$$A = \begin{bmatrix} 45 & 68 & 96 \\ 65 & 66 & 34 \\ 71 & 16 & 59 \\ 75 & 12 & 22 \\ 28 & 50 & 75 \end{bmatrix} \quad A_z = \begin{bmatrix} -0.5958 & 0.9522 & 1.2928 \\ 0.4141 & 0.8778 & -0.7730 \\ 0.7170 & -0.9820 & 0.0600 \\ 0.9190 & -1.1307 & -1.1729 \\ -1.4542 & 0.2827 & 0.5931 \end{bmatrix} \quad V_z = \begin{bmatrix} 0.6116 & -0.2317 & 0.7565 \\ -0.5309 & -0.8291 & 0.1753 \\ -0.5867 & 0.5088 & 0.6301 \end{bmatrix}$$

$$Uz = \begin{bmatrix} -0.5454 & 0.0043 & 0.5401 & 0.6398 & -0.0373 \\ 0.0806 & -0.8356 & -0.0203 & 0.1224 & 0.5291 \\ 0.3097 & 0.4659 & 0.4154 & -0.0481 & 0.7157 \\ 0.6198 & 0.0878 & -0.2462 & 0.7278 & -0.1336 \\ -0.4647 & 0.2775 & -0.6890 & 0.2089 & 0.4344 \end{bmatrix} \quad \Sigma_z = \begin{bmatrix} 2.9855 & 0 & 0 \\ 0 & 1.4566 & 0 \\ 0 & 0 & 0.9825 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Such that $A_z = U_z \Sigma_z V_z^* = U_z \Sigma_z V_z^T$.

1. Which documents treat the most similar topics?
2. Which terms are the most semantically similar?

Example

Solution:
In the context of LSA, both questions are similar in nature (although the arithmetics are of course different). In both cases, we have to "go back" on our steps and reconstruct the spaces of similarity.

You may remember from Wk 3 where we very briefly presented cMDS as a duality with PCA but with the particular difference of departing from pair-wise distances i.e. similarities, instead of coordinates as PCA does.

So if you have to build a rationale for solving this problem, how would you build a pair-wise similarity matrix out of the information of SVD analogous to cMDS?

Latent Semantic Analysis

Example

Solution (Cont.):

We know that cMDS just like PCA is an eigendecompotion, so let me rephrase the previous question; which eigendecomposition can you *undo* in order to get the documents similarity matrix D?.

And analogously, which eigendecomposition can you *undo* in order to get the terms similarity matrix W?

Note that we have studied both even if we did not gave them any name!

1. $D = V_z \Sigma_z^2 V_z * = A_z^T A_z$
2. $W = U_z \Sigma_z^2 U_z * = A_z A_z^T$

Latent Semantic Analysis

Example

Solution (Cont.):

So we start by calculating both D and W and note that you can calculate each one either using the eigendeomposition or "directly" using the product of the matrix of occurrences with its transpose. Regardless of how you choose to calculate it[a]:

$$D = V\Sigma^2 V* = A^T A = \begin{bmatrix} 4.0000 & -2.3582 & -2.9878 \\ -2.3582 & 4.0000 & 1.9875 \\ -2.9878 & 1.9875 & 4.0000 \end{bmatrix}$$

$$W = U\Sigma^2 U* = AA^T = \begin{bmatrix} 2.9331 & -0.4103 & -1.2847 & -3.1406 & 1.9025 \\ -0.4103 & 1.5396 & -0.6115 & 0.2946 & -0.8125 \\ -1.2847 & -0.6115 & 1.4820 & 1.6990 & -1.2847 \\ -3.1406 & 0.2946 & 1.6990 & 3.4988 & -2.3517 \\ 1.9025 & -0.8125 & -1.2847 & -2.3517 & 2.5465 \end{bmatrix}$$

[a] If you opt for using the eigendecomposition do not forget to "remove" the corresponding null subspaces as appropriate to make matrices comformable for the product.

Example

Solution (Cont.):

The values in D and W represent similarity among the entities. The values themsolves have no absolute meaning; they only have meaning as relative to each other.

- Example; It does not matter that the similarity of $d1$ and $d2$ is $D(1,2) = -2.3582$, what matters is whether this value is bigger or smaller than other entries in D.

Latent Semantic Analysis

Example

Solution (Cont.):

We are now finally ready to answer both questions:

- Which documents treat the most similar topics?

  The magnitude of values in D are *proportional*[a] to the similarity between the document vectors, with higher values in D representing more similarity between the documents compared.

  Hence in this case, documents $d1$ and $d3$ are the most similar i.e. $|D(1,3)| = 2.98$

- Which terms are the most semantically similar?

  Analogously, higher magnitudes represent more semantically similar terms, ergo the terms $t1$ and $t4$ are the most similar i.e. $|W(1,4)| = 3.14$

[a] The magnitude of values in D are *proportional* to the cosine of the angle between the corresponding document vectors in the reduced-dimensional semantic space. The sign represents whether the rotation is clock-wise or anti-clockwise.

Thanks

Thank you! Questions?