Introduction to Algorithmic Data Analysis

Esther Galbrun

Autumn 2023

UNIVERSITY OF
EASTERN FINLAND

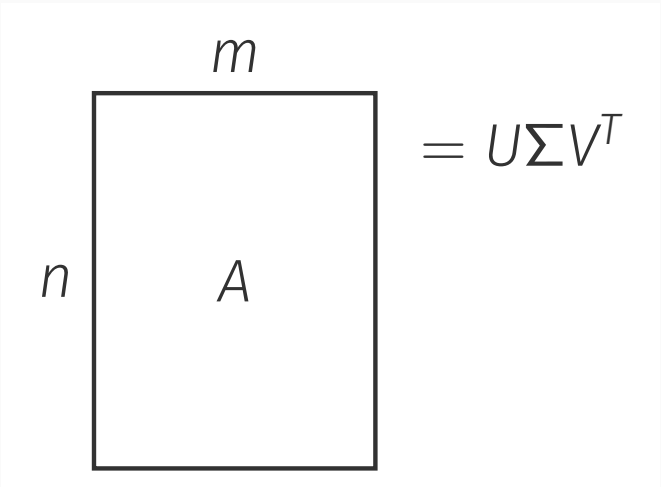
Last updated on 2023-07-04 10:09

Q0.1: Matrix sizes

Consider a matrix A and
its rank-k truncated singular value decomposition (SVD) $U\Sigma V^T$

A has size $n \times m$ Σ has size $k \times k$

i) What is the size of U? ii) What is the size of V?



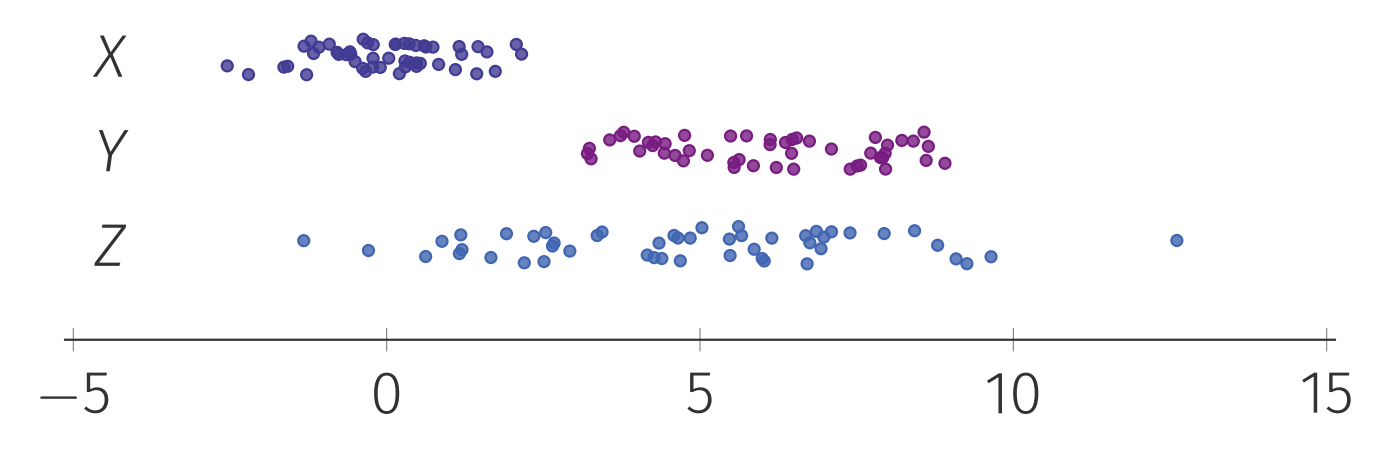
Q0.2: Point clouds (i)

Consider the three collections of points below

i) Which one has the largest median?

ii) Which one has the largest mean?

iii) Which one has the largest variance?



Q0.3: Point clouds (ii)

Consider the three collections of points below

one is sampled from a uniform distribution on $[3, 9]$, i.e. $\mathcal{U}(3, 9)$

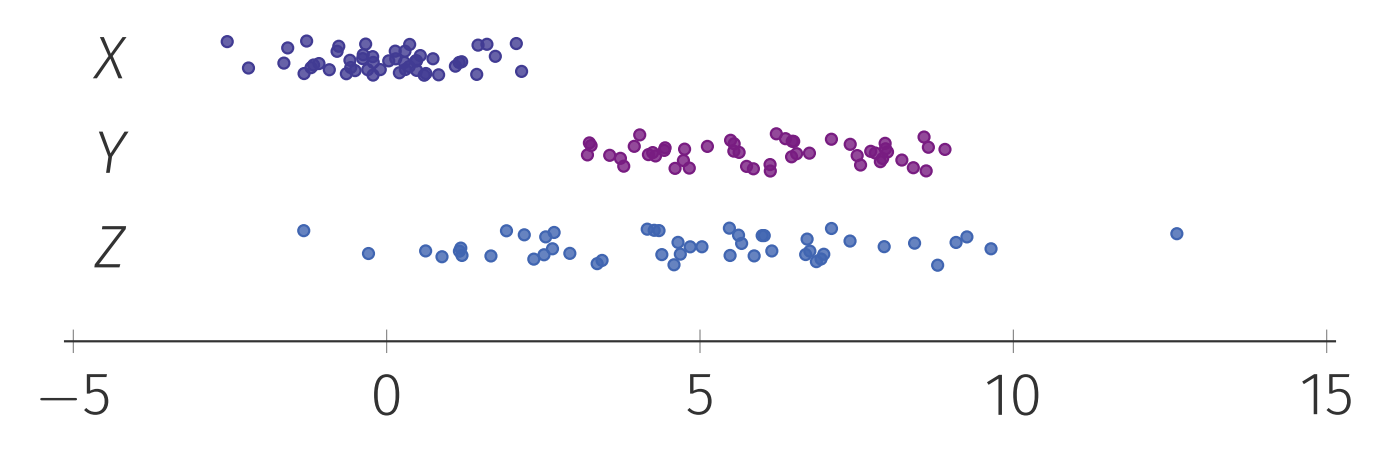
one from a Gaussian distribution

with mean $\mu = 0$ and variance $\sigma^2 = 1$, i.e. $\mathcal{N}(0, 1)$

one from a Gaussian distribution

with mean $\mu = 5$ and variance $\sigma^2 = 9$, i.e. $\mathcal{N}(5, 9)$

Which one is which?



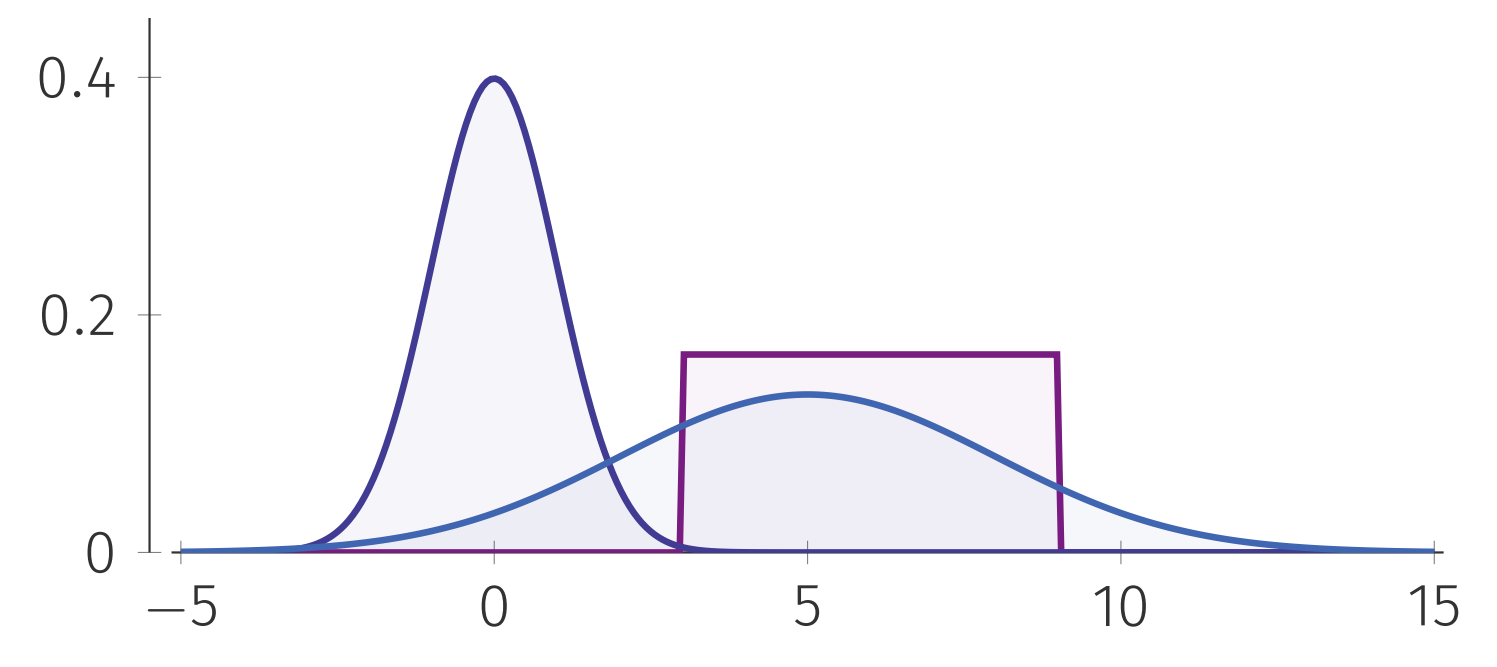
Q0.4: Bells and bricks

What is $P(x \leq 5)$?

i) Assuming $P \sim \mathcal{U}(3, 9)$

ii) Assuming $P \sim \mathcal{N}(0, 1)$

iii) Assuming $P \sim \mathcal{N}(5, 9)$



Q0.5: Counting letters

Consider the following sentence:

Kaikki ihmiset syntyvät vapaina ja tasavertaisina arvoltaan ja oikeuksiltaan. Heille on annettu järki ja omatunto, ja heidän on toimittava toisiaan kohtaan veljeyden hengessä.

Fill in the contingency table

	a	$\bar{a}$
y		
$\bar{y}$		

a word contains ‘a’

$\bar{a}$ word does not contain ‘a’

y word contains ‘y’

$\bar{y}$ word does not contain ‘y’

Q0.6: Co-occurring letters

	a	$\bar{a}$	
y	0	2	2
$\bar{y}$	14	7	21
	14	9	23

a : word contains ‘a’

$\bar{a}$: word does not contain ‘a’

y : word contains ‘y’

$\bar{y}$: word does not contain ‘y’

i) What is $P(a \wedge y)$ estimated from the counts?

ii) What is $P(a \wedge y)$ estimated under the assumption that a and y are independent?

Q0.7: Marbles (i)

A bag contains 10 marbles, 2 of which are red.
The event that we draw a red marble constitutes a success.
We draw 3 times,
and denote as $Y_i \in \{0, 1\}$ the outcome of the i^{th} draw.

Is it more less likely that the third draw is a success,
knowing that the first two draws failed?

$$P(Y_3 = 1 \mid Y_1 = 0, Y_2 = 0) \overset{?}{\lessgtr} P(Y_3 = 1)$$

Q0.8: Marbles (ii)

A bag contains 10 marbles, 2 of which are red.
The event that we draw a red marble constitutes a success.
We draw 3 times, and denote as X the number of successes.

Compute $P(X = k)$ for $k \in \{0, 1, 2, 3\}$
both with and without replacement

Q0.9: Mystery value

Given the multiset $X = \{1, 2, 4, 7, 9, 12, 14\}$.

What can you say about $x \in \mathbb{N}$ if you know ...

i) $\textsf{mean}(X \cup \{x\}) = 8$?

ii) $\textsf{median}(X \cup \{x\}) = 8$?

iii) $\textsf{mean}(X \cup \{x\}) = 10$?

iv) $\textsf{median}(X \cup \{x\}) = 10$?

v) $\textsf{mean}(X \cup \{x\}) = 6.5$?

vi) $\textsf{median}(X \cup \{x\}) = 6.5$?