Week 7: Clustering and Dimensionality Reduction

Matthew Caldwell

COMP0088 Introduction to Machine Learning • UCL Computer Science • Autumn 2024

Admin

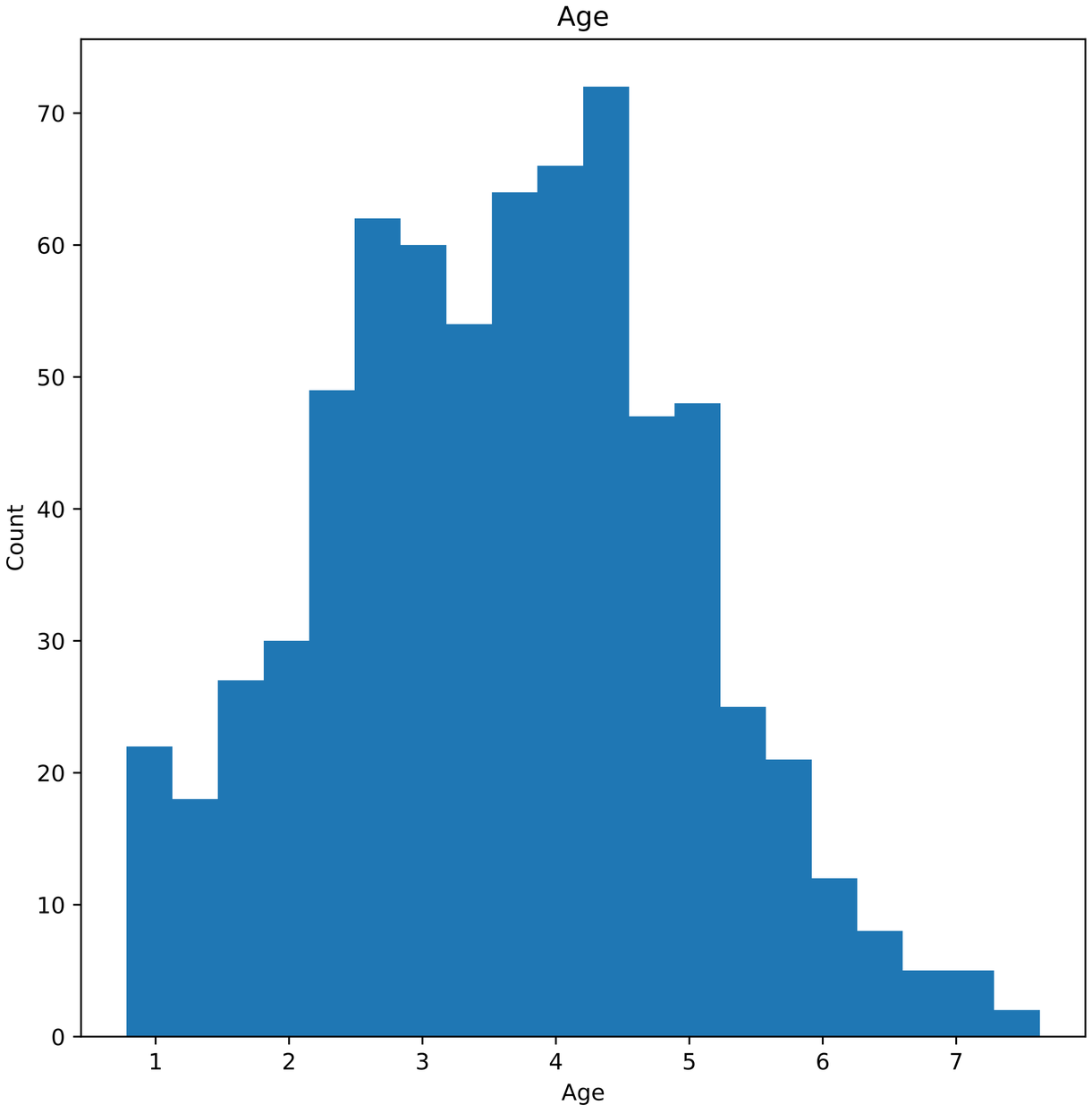
- Pulse surveys feedback

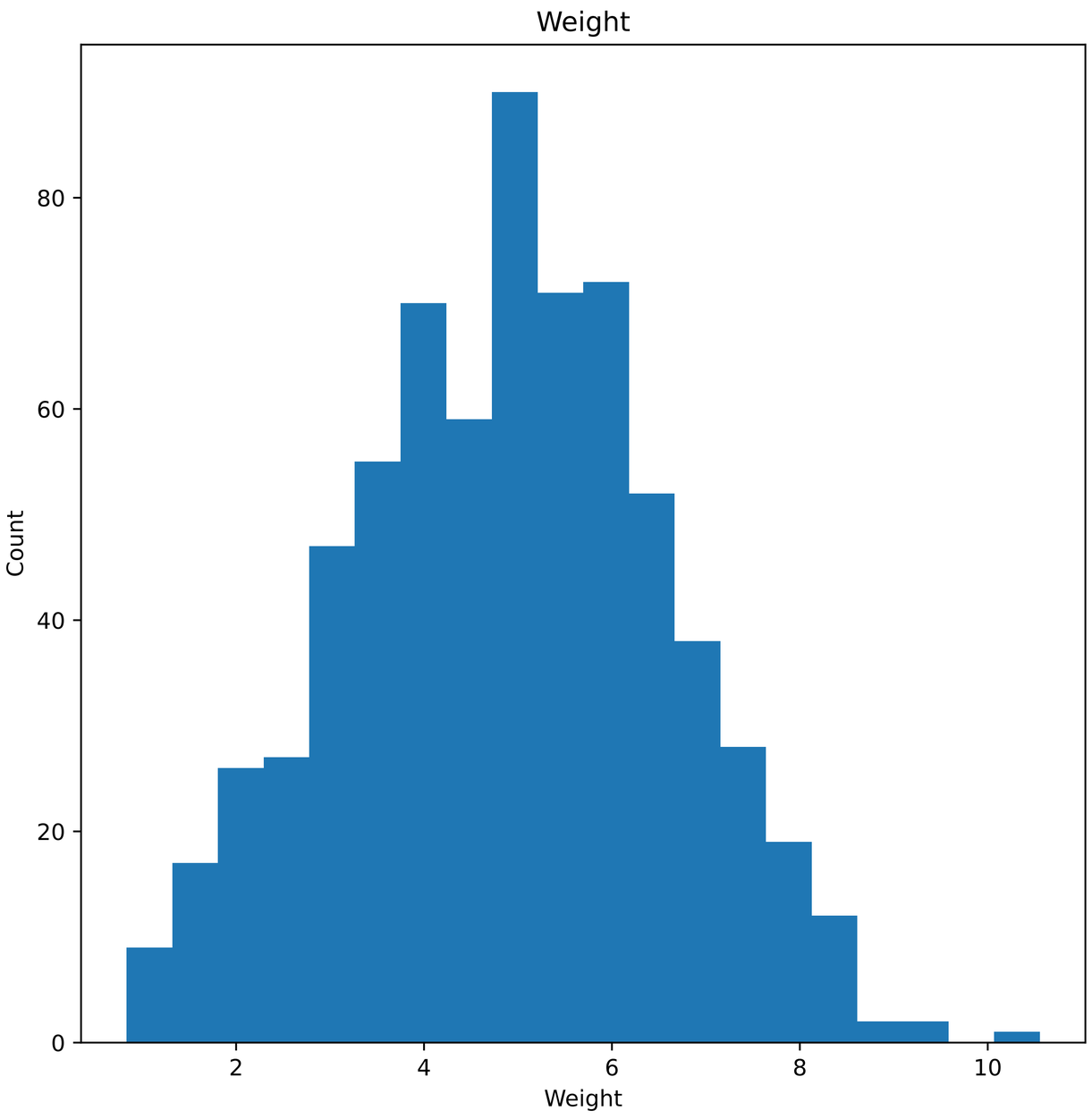
Week 7 Recap

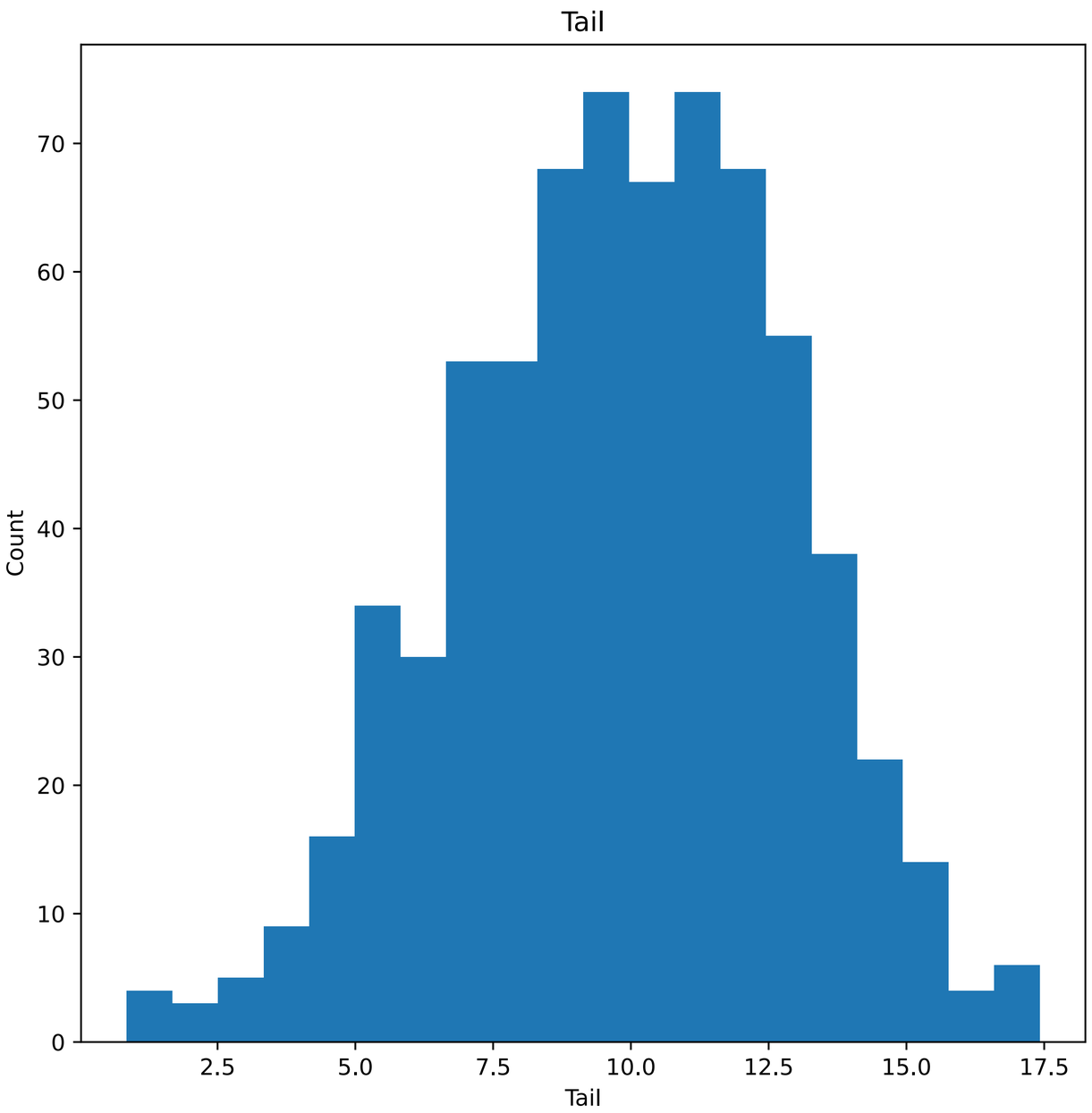
Groundhog Day

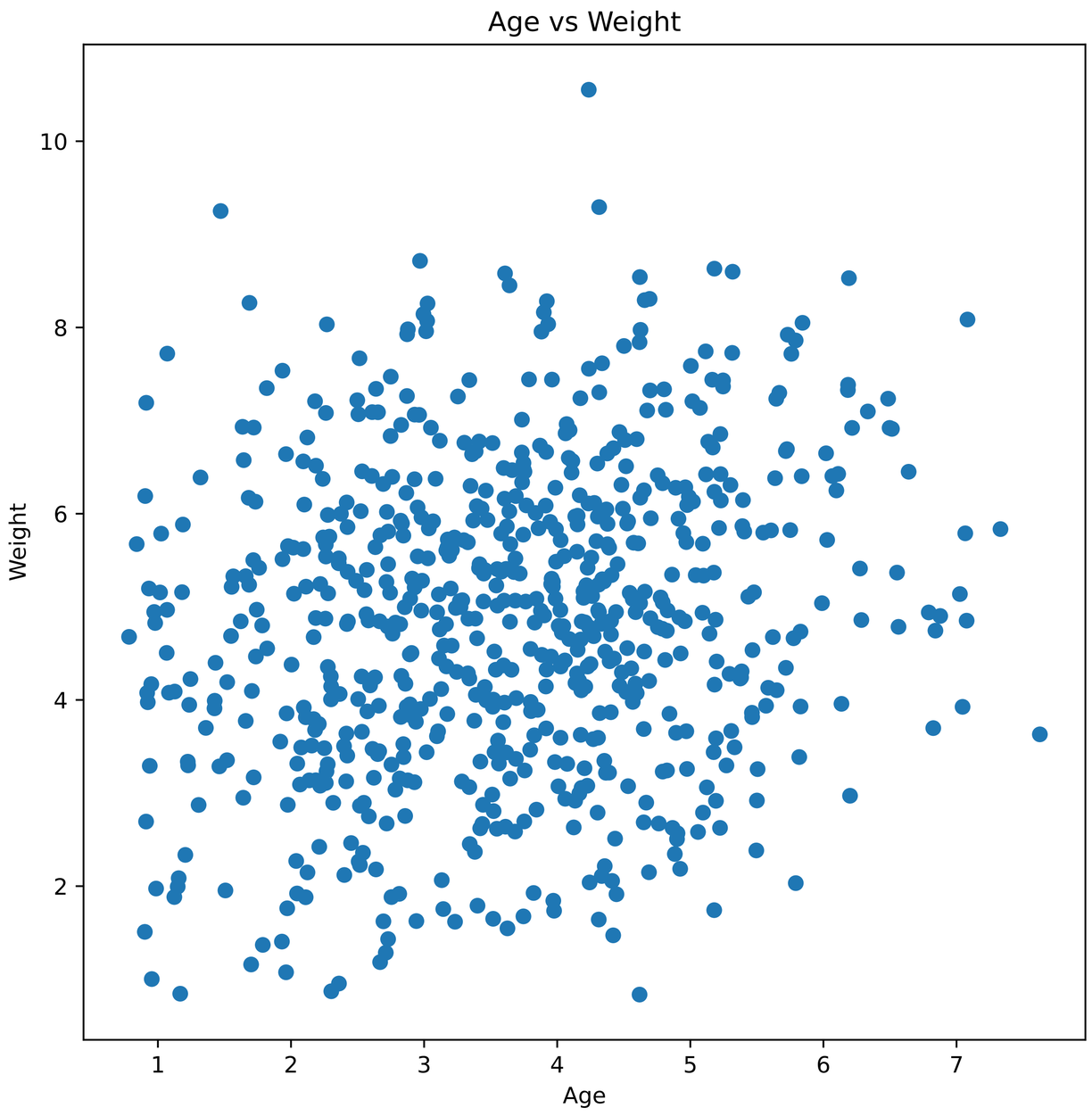
groundhogs

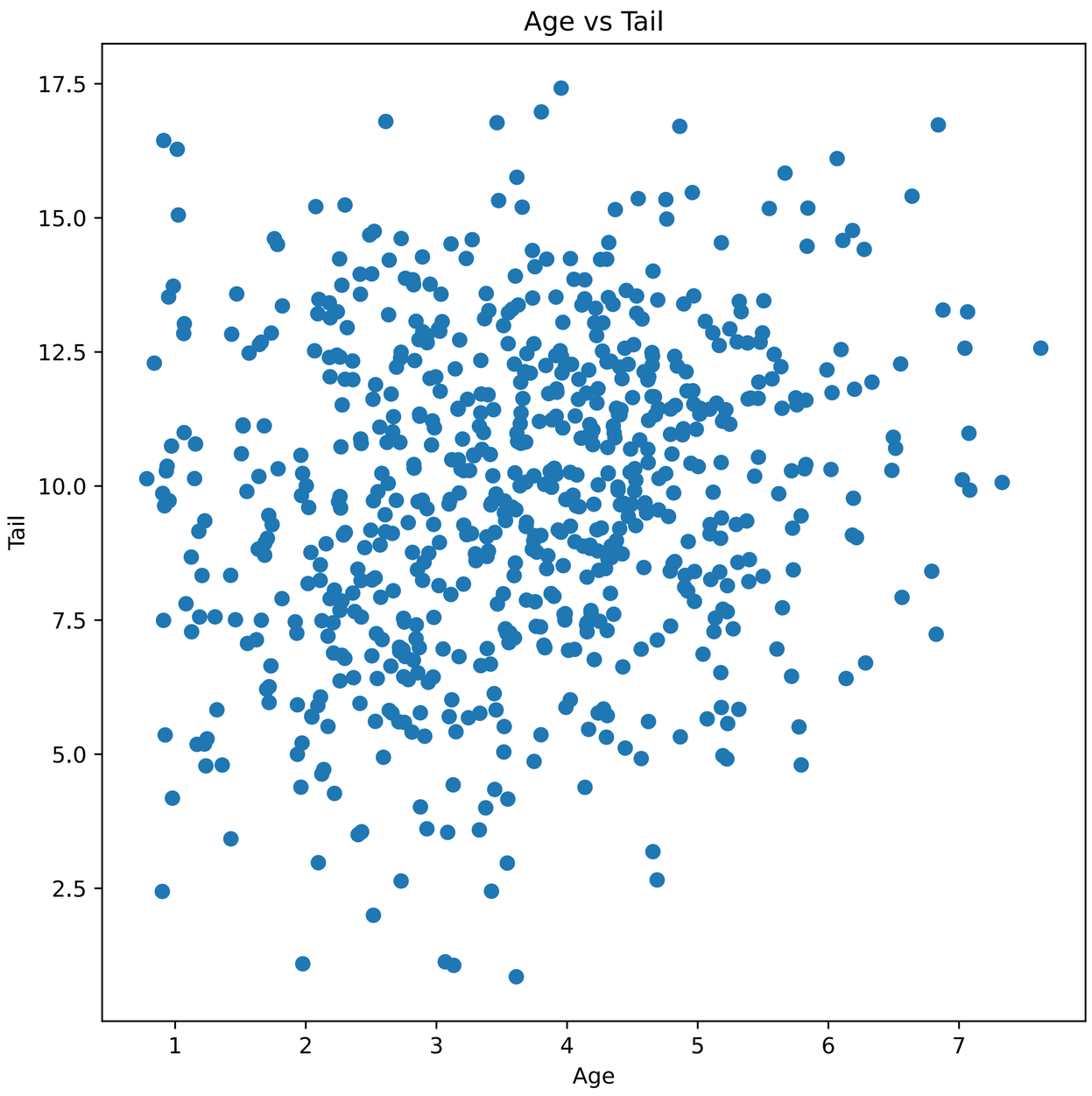
	Age	Weight	Height	Length	Tail	Whiskers	Incisors	E
0	2.2	3.8	14.3	14.7	5.5	4.5	1.4	
1	4.9	2.6	11.8	13.3	8.3	0.9	2.2	
2	2.6	5.2	12.6	15.9	9.9	1.9	3.6	
3	7.5	4.3	7.0	15.2	9.3	0.5	1.8	
4	2.2	3.1	11.4	15.0	12.0	1.8	2.6	
5	5.2	6.1	13.2	15.7	5.6	3.3	3.3	
6	3.5	3.1	7.7	20.2	7.4	0.7	2.9	
7	3.7	1.7	2.8	15.3	12.7	3.2	3.8	
8	2.5	2.4	9.6	14.1	7.2	1.5	2.0	
9	4.4	6.6	5.4	20.9	8.8	2.6	4.0	
10	3.3	5.7	9.1	18.2	3.6	1.7	3.6	
11	4.3	4.7	9.4	19.6	7.1	3.5	-0.4	
12	7.3	5.8	12.3	12.9	10.1	3.8	2.4	
	[illegible]	4.7	6.5	16.4	[illegible]	[illegible]	4.6	

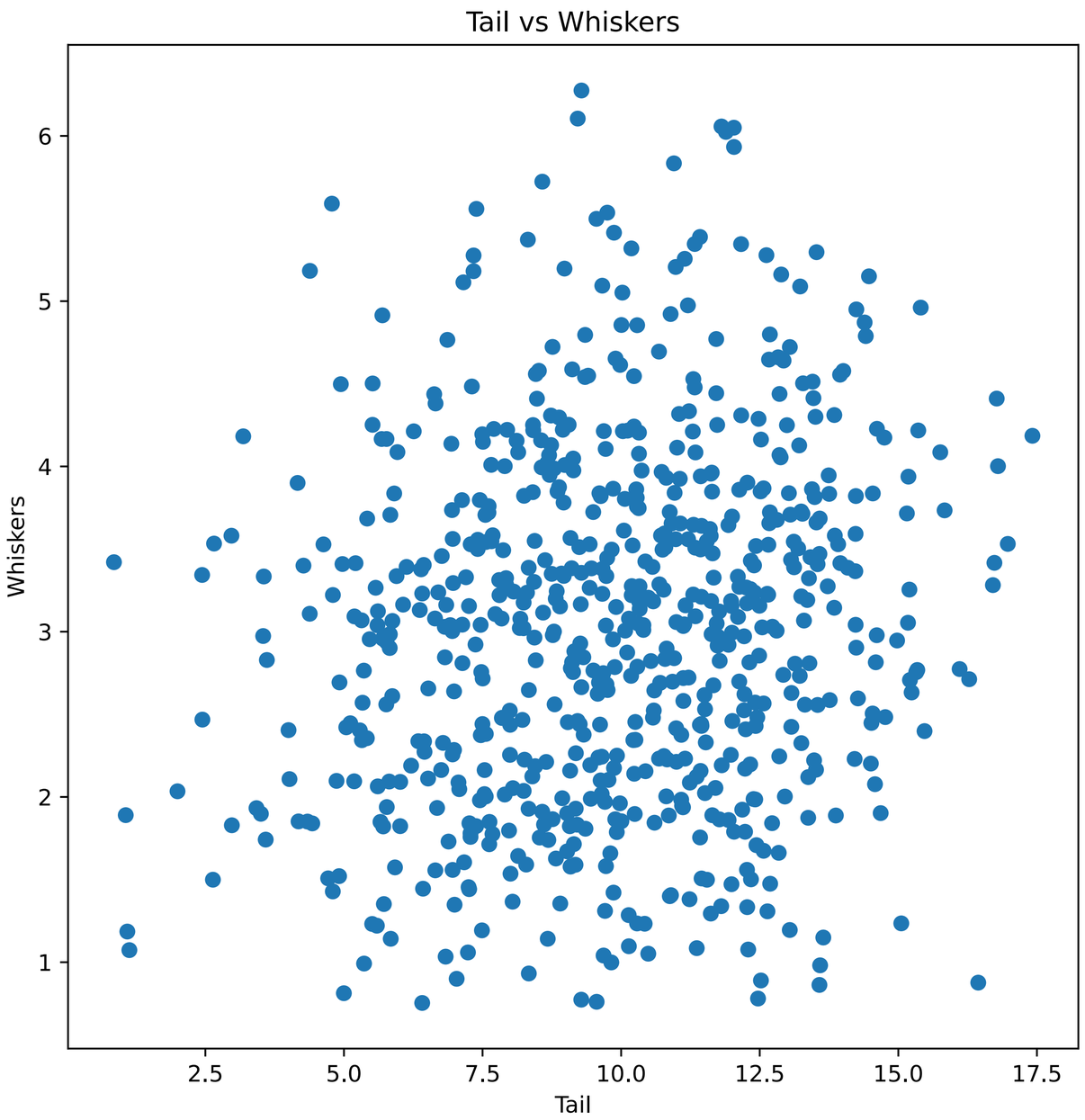

Age
Count
Age
0
10
20
30
40
50
60
70
1
2
3
4
5
6
7

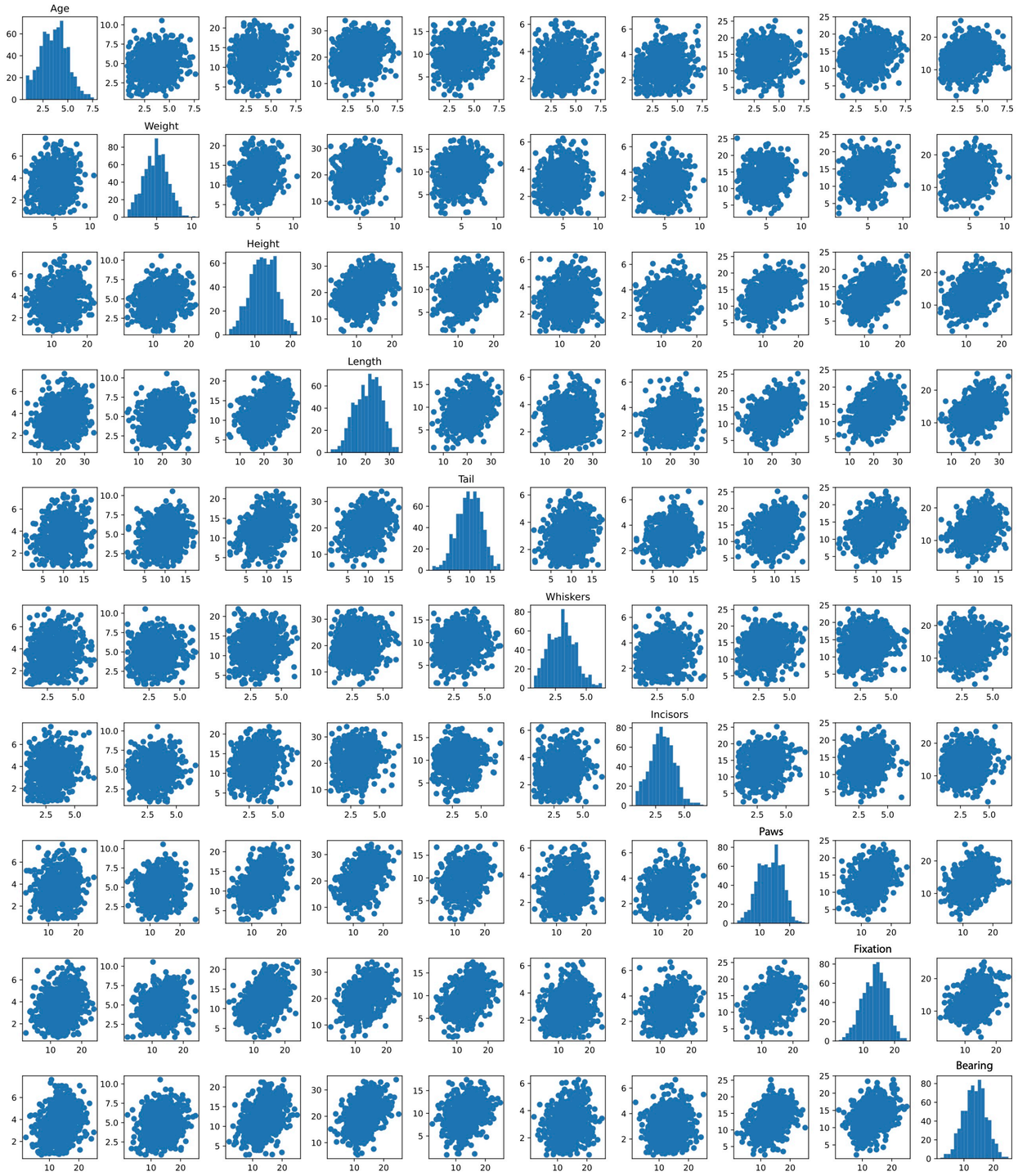

Weight
Count
0
20
40
60
80
2
4
6
8
10
Weight


Tail
Count
0
10
20
30
40
50
60
70
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Tail


Age vs Weight
Weight
Age
10
8
6
4
2
1
2
3
4
5
6
7

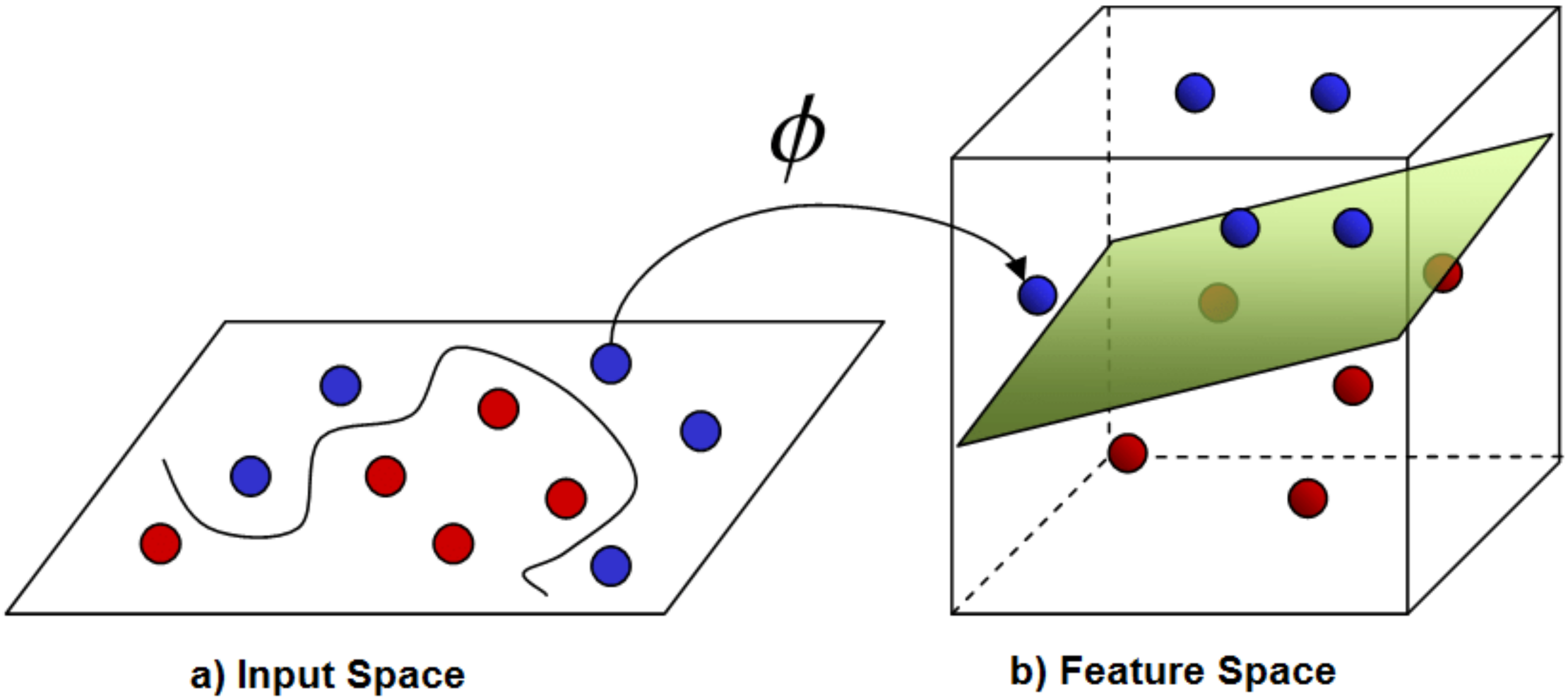

Age vs Tail
17.5
15.0
12.5
10.0
7.5
5.0
2.5
Tail
1
2
3
4
5
6
7
Age

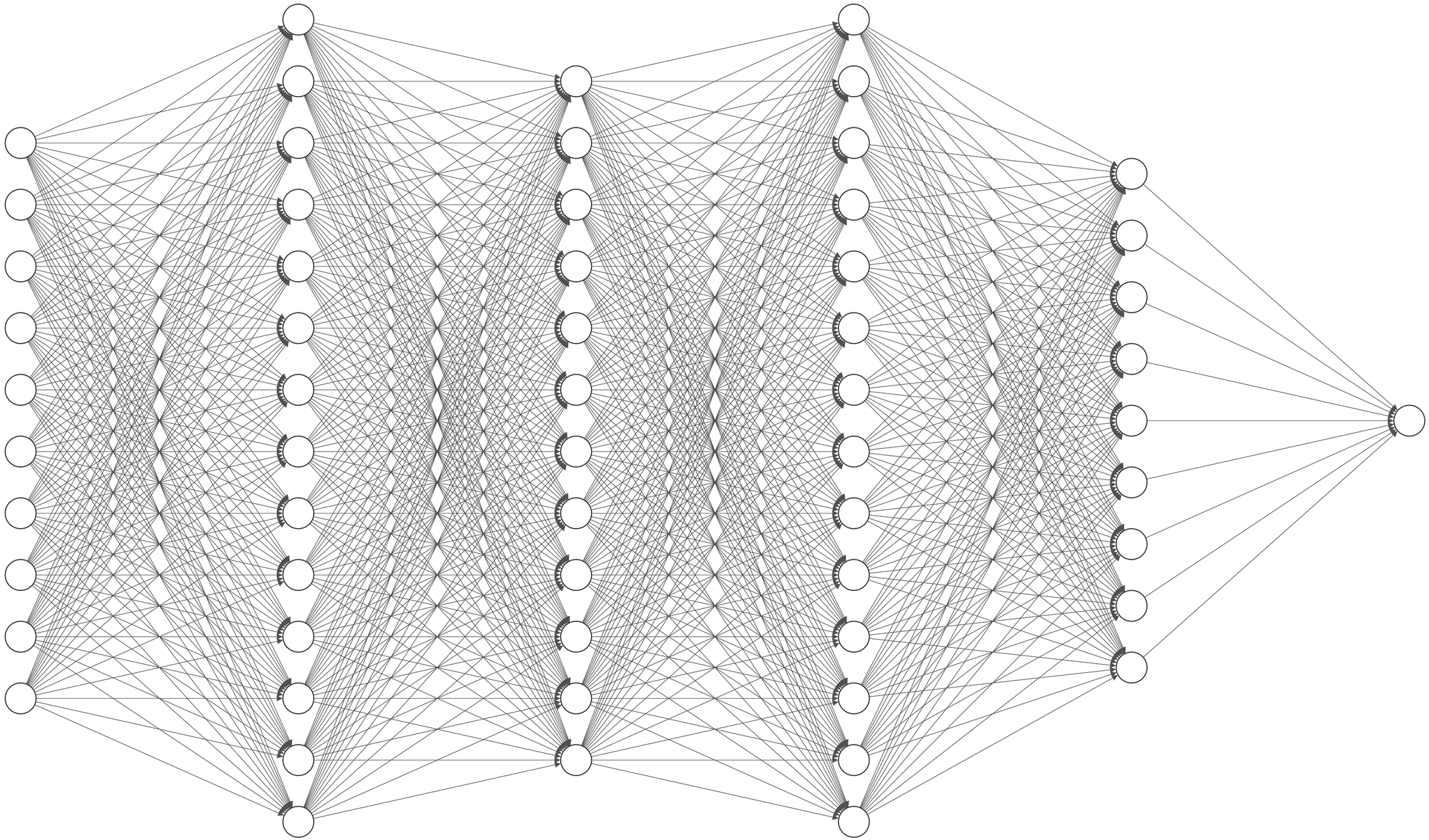

Tail vs Whiskers
Whiskers
Tail
1
2
3
4
5
6
2.5
5.0
7.5
10.0
12.5
15.0
17.5

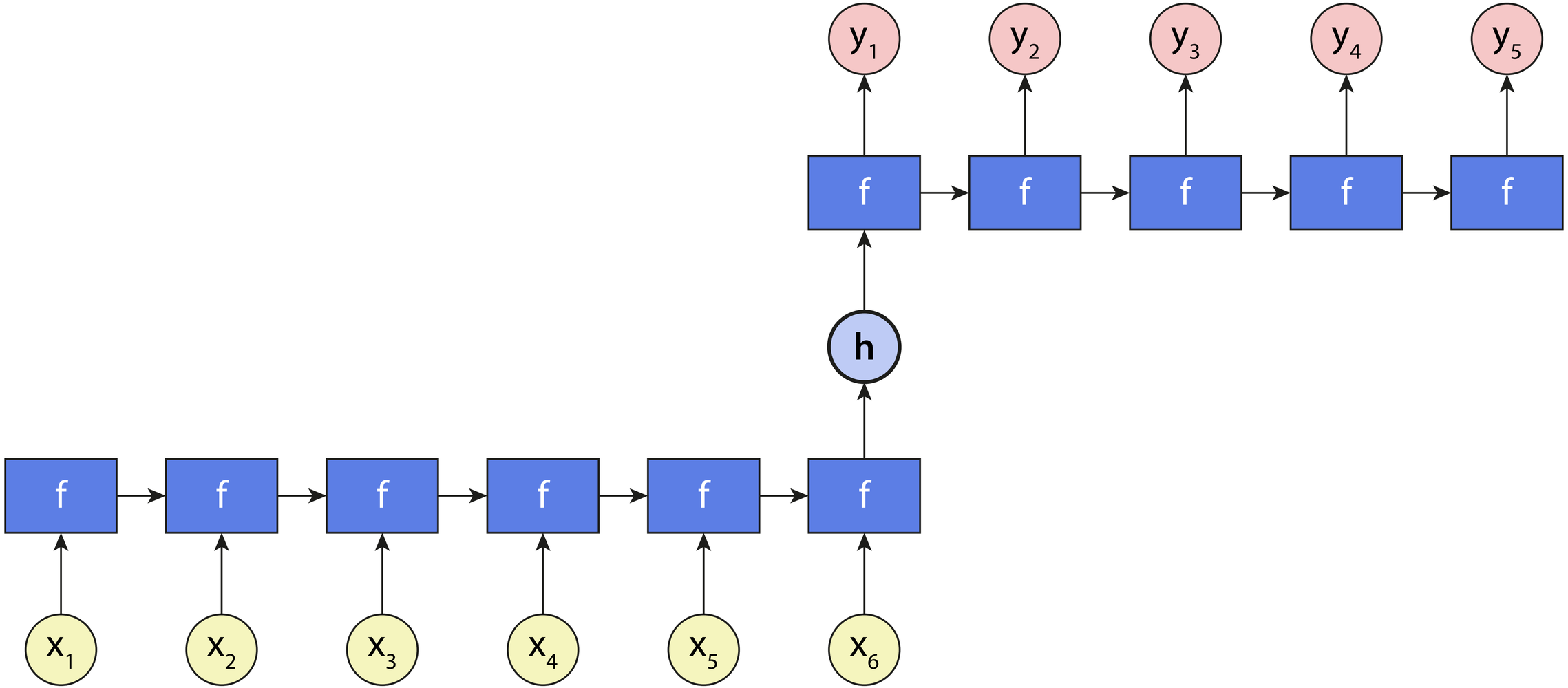

Age
Weight
Height
Length
Tail
Whiskers
Incisors
Paws
Fixation
Bearing

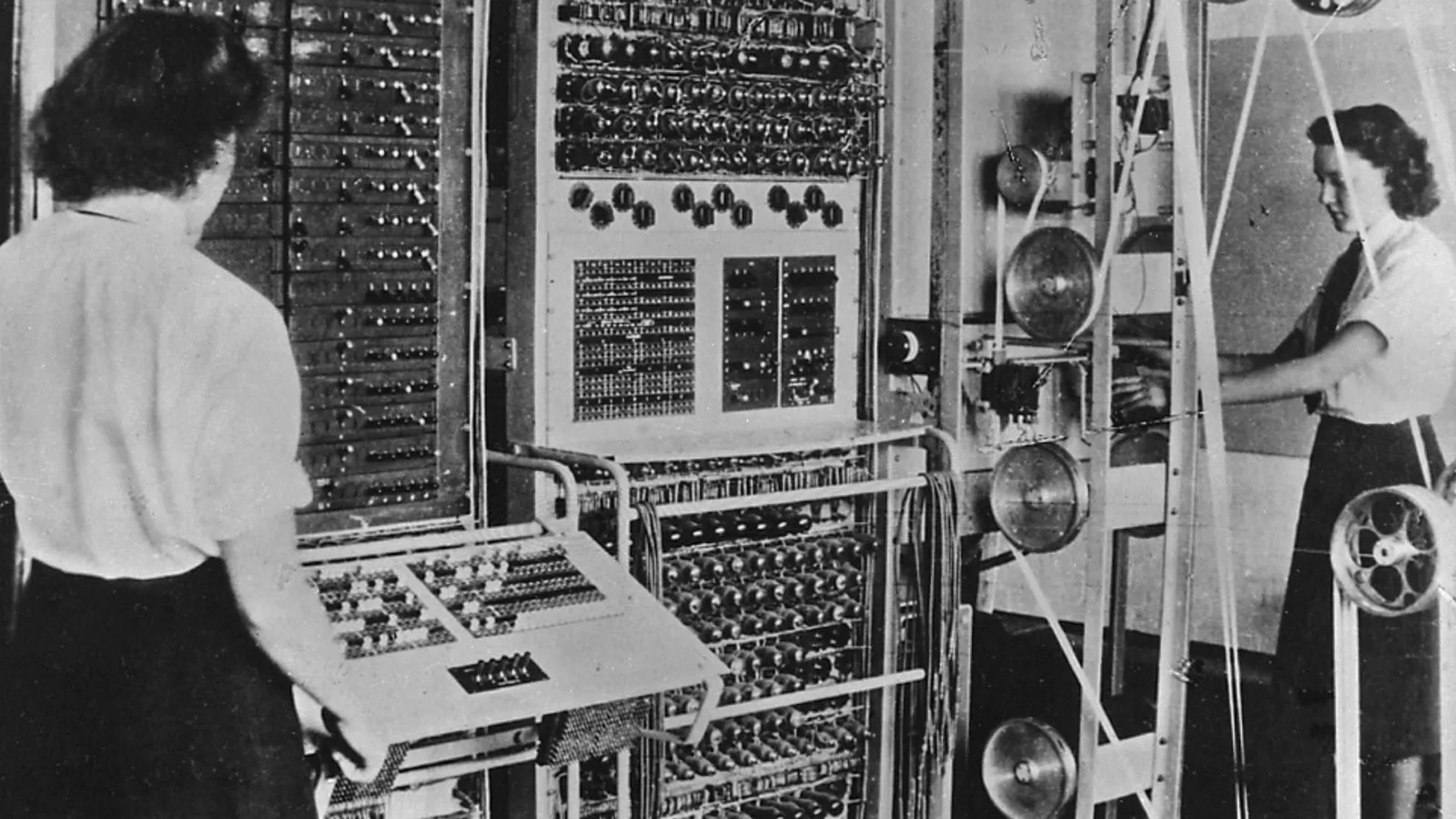
Mayer

embeddings


ϕ
a) Input Space
b) Feature Space




y1
y2
y3
y4
y5
f
f
f
f
f
h
f
f
f
f
f
f
x1
x2
x3
x4
x5
x6



a
ability
able
about
above
accept
according
account
across
act
action
activity
actually
add
address
administration
admit
adult
affect
after
again
against
age
agency
agent
ago
agree
agreement
ahead
air
between
beyond
big
bill
billion
bit
black
blood
blue
board
body
book
born
both
box
boy
break
bring
brother
budget
build
building
business
but
buy
by
call
camera
campaign
can
doctor
dog
door
down
draw
dream
drive
drop
drug
during
each
early
east
easy
eat
economic
economy
edge
education
effect
effort
eight
either
election
else
employee
end
energy
enjoy
enough
husband
I
idea
identify
if
image
imagine
impact
important
improve
in
include
including
increase
indeed
indicate
individual
industry
information
inside
instead
institution
interest
interesting
international
interview
into
investment
involve
issue
material
matter
may
maybe
me
mean
measure
media
medical
meet
meeting
member
memory
mention
message
method
middle
might
military
million
mind
minute
miss
mission
model
modern
moment
money
month
more
oil
ok
old
on
once
one
only
onto
open
operation
opportunity
option
or
order
organization
other
others
our
out
outside
over
own
owner
page
pain
painting
paper
parent
part
participant
perhaps
period
person
personal
phone
physical
pick
picture
piece
place
plan
plant
play
player
PM
point
police
policy
political
politics
poor
popular
population
position
positive
possible
power
practice
prepare
present
really
reason
receive
recent
recently
recognize
record
red
reduce
reflect
region
relate
relationship
religious
remain
remember
remove
report
represent
Republican
require
research
resource
respond
response
responsibility
rest
result
return
reveal
shoulder
show
side
sign
significant
similar
simple
simply
since
sing
single
sister
sit
site
situation
six
size
skill
skin
small
smile
so
social
society
soldier
some
somebody
someone
something
sometimes
understand
unit
until
up
upon
us
use
usually
value
various
very
victim
view
violence
visit
voice
vote
wait
walk
wall
want
war
watch
water
way
we
weapon
wear
week
weight

$\mathbf{x}^T$

(one-hot)

$$\mathbf{x}^\mathsf{T} \longrightarrow \tilde{\mathbf{x}}^\mathsf{T}$$

(one-hot) (embedding)

$\mathbf{x}^\top$

(one-hot)

$\mathbf{W}$

$\tilde{\mathbf{x}}^\top$

(embedding)

$\mathbf{x}^T$

(one-hot)

$\mathbf{W}$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$\tilde{\mathbf{x}}^T$

(embedding)

$\mathbf{x}^T$

(one-hot)

[0, 1, 0, 0, 0]

$\mathbf{W}$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$\tilde{\mathbf{x}}^T$

(embedding)

$\mathbf{x}^T$

(one-hot)

$[\ 0, 1, 0, 0, 0\ ] \quad \times$

$\mathbf{W}$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$\tilde{\mathbf{x}}^T$

(embedding)

$\mathbf{x}^T$

(one-hot)

$\mathbf{W}$

$\tilde{\mathbf{x}}^T$

(embedding)

$[\ 0, 1, 0, 0, 0\ ] \times$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$= [\qquad\qquad\qquad

$\mathbf{x}^T$

(one-hot)

$\mathbf{W}$

$\tilde{\mathbf{x}}^T$

(embedding)

$[\ 0, 1, 0, 0, 0\ ] \times$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$= [\qquad\qquad\qquad

$\mathbf{x}^T$

(one-hot)

$\mathbf{W}$

$\tilde{\mathbf{x}}^T$

(embedding)

$[\,0, 1, 0, 0, 0\,] \times$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$= [\,0.45 \qquad\qquad\qquad\

$\mathbf{x}^T$

(one-hot)

$\mathbf{W}$

$\tilde{\mathbf{x}}^T$

(embedding)

$[\,0, 1, 0, 0, 0\,] \times$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$= [\,0.45 \qquad\qquad\qquad\

$\mathbf{x}^T$

(one-hot)

$\mathbf{W}$

$\tilde{\mathbf{x}}^T$

(embedding)

$[\,0, 1, 0, 0, 0\,] \times$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$= [\,0.45, 0.03\,]$

$\mathbf{x}^T$

(one-hot)

$\mathbf{W}$

$\tilde{\mathbf{x}}^T$

(embedding)

$[\,0, 1, 0, 0, 0\,] \times$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$= [\,0.45, 0.03 \quad\quad\quad\

$\mathbf{x}^T$

(one-hot)

$\mathbf{W}$

$\tilde{\mathbf{x}}^T$

(embedding)

$[\ 0, 1, 0, 0, 0\ ] \times$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$= [\ 0.45, 0.03, 0.88\ ]$

$\mathbf{x}^T$

(one-hot)

$\mathbf{W}$

$\tilde{\mathbf{x}}^T$

(embedding)

$[\ 0, 1, 0, 0, 0\ ] \times$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$= [\ 0.45, 0.03, 0.88\ ]$

$\mathbf{x}^T$

(one-hot)

$\mathbf{W}$

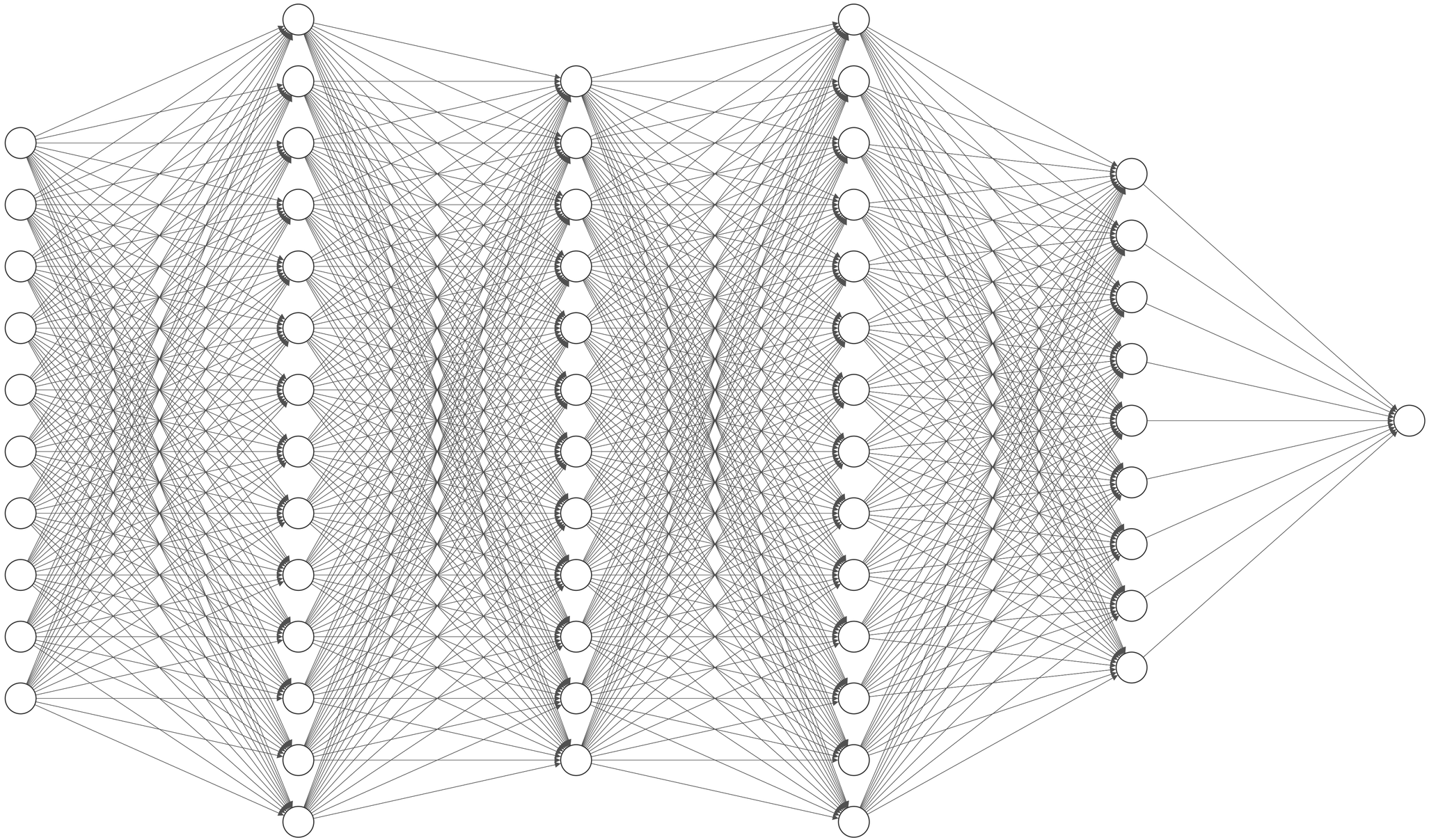
$\tilde{\mathbf{x}}^T$

(embedding)

$[\ 0, 1, 0, 0, 0\ ] \times$

	e_1	e_2	e_3
vampire	0.99	0.64	0.15
groundhogs	0.45	0.03	0.88
cast	0.02	0.46	0.6
no	0.73	0.89	0.9
shadow	0.83	0.75	0.44

$= [\ 0.45, 0.03, 0.88\ ]$



In the population of Transylvania there are four distinct nationalities: Saxons in the South, and mixed with them the Wallachs, who are the descendants of the Dacians; Magyars in the West, and Szekelys in the East and North. I am going among the latter, who claim to be descended from Attila and the Huns. This may be so, for when the Magyars conquered the country in the eleventh century they found the Huns settled in it. I read that every known superstition in the world is gathered into the horseshoe of the Carpathians, as if it were the centre of some sort of imaginative whirlpool; if so my stay may be very interesting. (*Mem.*, I must ask the Count all about them.)

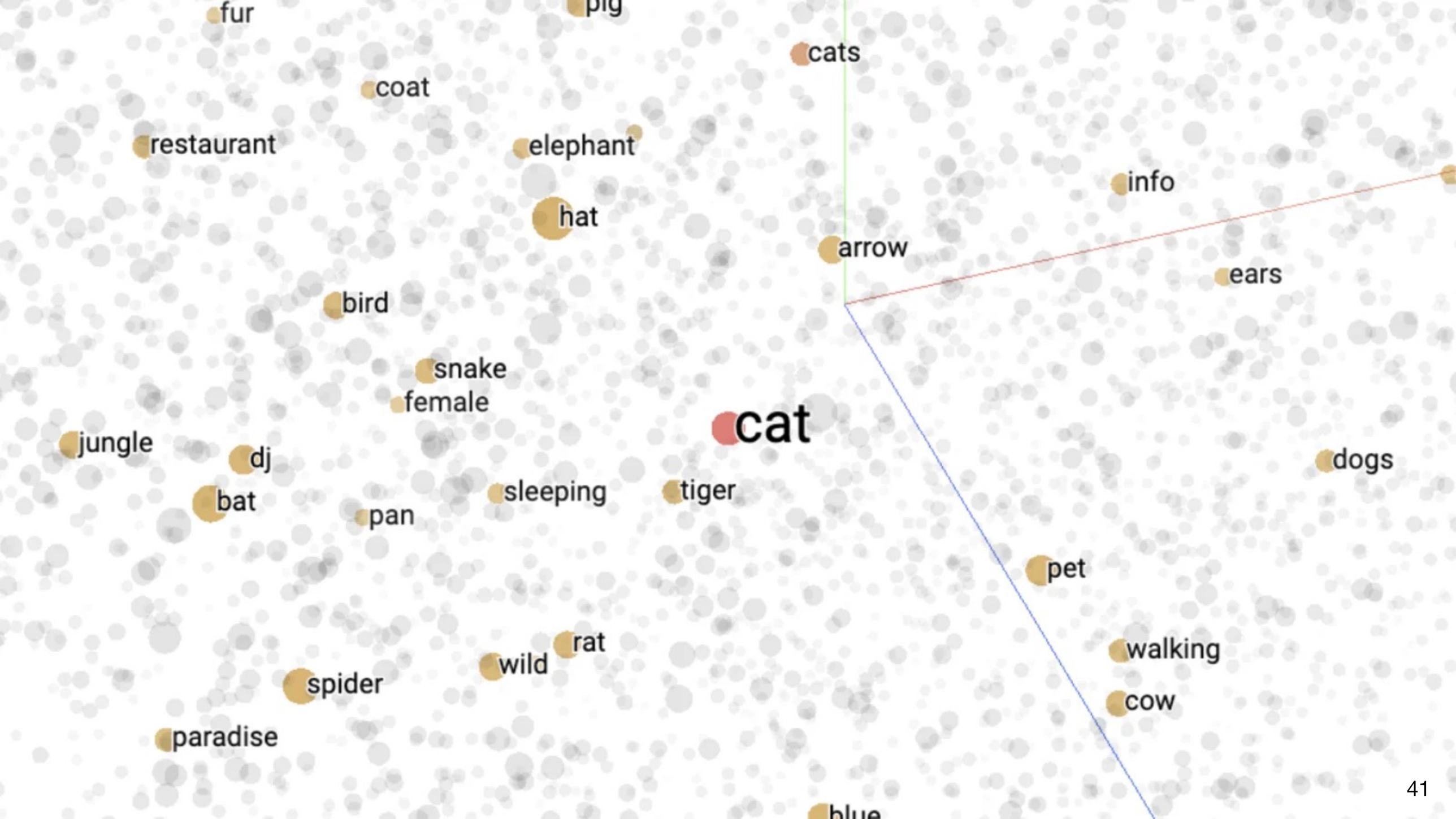
I did not sleep well, though my bed was comfortable enough, for I had all sorts of queer dreams. There was a dog howling all night under my window, which may have had something to do with it; or it may have been the paprika, for I had to drink up all the water in my carafe, and was still thirsty. Towards morning I slept and was wakened by the continuous knocking at my door, so I guess I must have been sleeping soundly then. I had for breakfast more paprika, and a sort of porridge of maize flour which they said was "mamaliga," and egg-plant stuffed with forcemeat, a very excellent dish, which they call "impletata." (*Mem.*, get recipe for this also.) I had to hurry breakfast, for the train started a little before eight, or rather it ought to have done so, for after rushing to the station at 7:30 I had to sit in the carriage for more than an hour before we began to move. It seems to me that the further east you go the more unpunctual are the trains. What ought they to be in China?

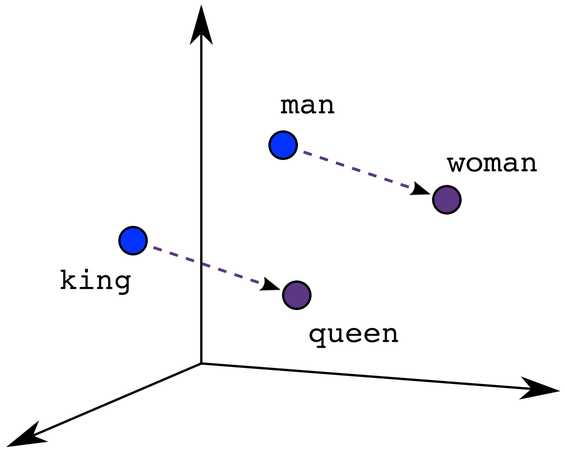
All day long we seemed to dawdle through a country which was full of beauty of every kind. Sometimes we saw little towns or castles on the top of steep hills such as we see in old missals; sometimes we ran by rivers and streams which seemed from the wide stony margin on each side of them to be subject to great floods. It takes a lot of water, and running strong, to sweep the outside edge of a river clear. At every station there were groups of people, sometimes crowds, and in all sorts of attire. Some of them were just like the peasants at home or those I saw coming through France and Germany, with short jackets and round hats and home-made trousers; but others were very picturesque. The women looked pretty, except when you got near them, but they were very clumsy about the waist. They had all full white sleeves of some kind or other, and most of them had big belts with a lot of strips of something fluttering from them like the dresses in a ballet, but of course there were petticoats under them. The strangest figures we saw were the Slovaks, who were more barbarian than the rest, with their big cow-boy hats, great baggy dirty-white trousers, white linen shirts, and enormous heavy leather belts, nearly a foot wide, all studded over with brass nails. They wore high boots, with their trousers tucked into them, and had long black hair and heavy black moustaches. They are very picturesque, but do not look prepossessing. On the stage they would be set down at once as some old Oriental band of brigands. They are, however, I am told, very harmless and rather wanting in natural self-assertion.

It was on the dark side of twilight when we got to Bistritz, which is a very interesting old place. Being practically on the frontier—for the Borgo Pass leads from it into Bukovina—it has had a very stormy existence, and it certainly shows marks of it. Fifty years ago a series of great fires took place, which made terrible havoc on five separate occasions. At the very beginning of the seventeenth century it underwent a siege of three weeks and lost 13,000 people, the casualties of war proper being assisted by famine and disease.

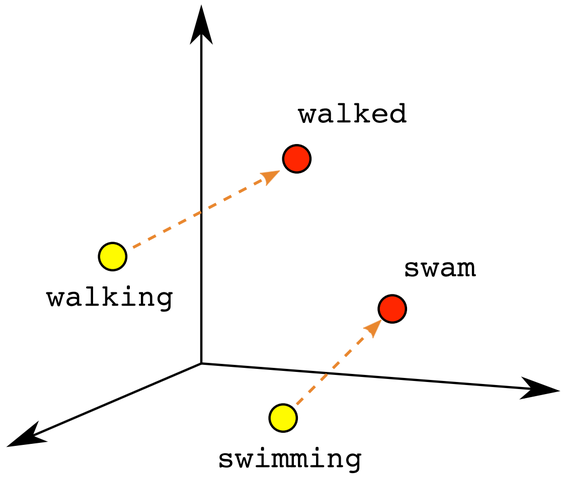
Count Dracula had directed me to go to the Golden Krone Hotel, which I found, to my great delight, to be thoroughly old-fashioned, for of course I wanted to see all I could of the ways of the country. I was evidently expected, for when I got near the door I faced a cheery-looking elderly woman in the usual peasant dress—white undergarment with long double apron, front, and back, of coloured stuff fitting almost too tight for modesty. When I came close she bowed and said, "The Herr Englishman?" "Yes," I said, "Jonathan Harker." She smiled, and gave some message to an elderly man in white shirt-sleeves, who had followed her to the door. He went, but immediately returned with a letter:—

"My Friend.—Welcome to the Carpathians. I am anxiously expecting you. Sleep well to-night. At three to-morrow the diligence will start for Bukovina; a place on it is kept for you. At the Borgo Pass my carriage will await you and will bring you to me. I trust that your journey from London has been a happy one, and that you will enjoy your stay in my beautiful land.

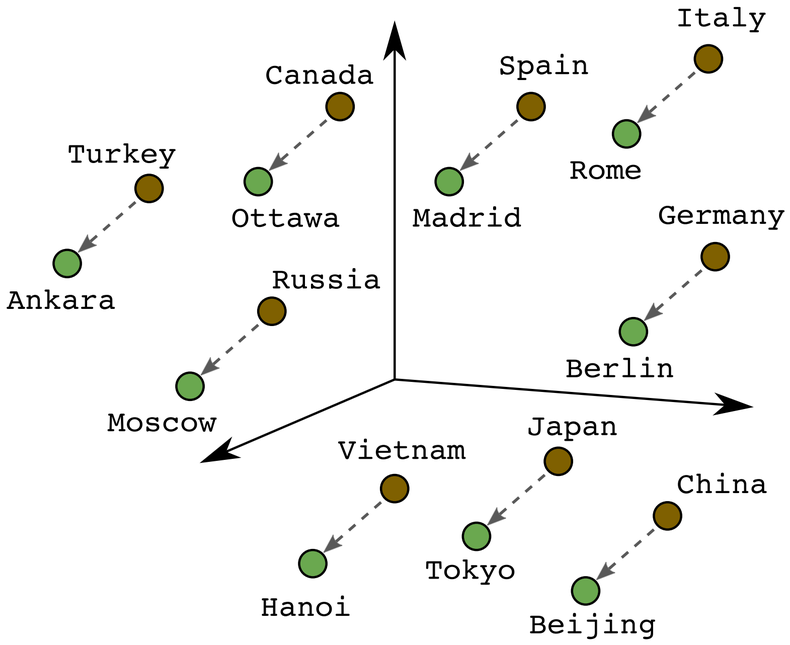

fur
pig
cats
coat
restaurant
elephant
info
hat
arrow
ears
bird
snake
female
cat
jungle
dj
dogs
bat
sleeping
tiger
pan
pet
rat
walking
wild
spider
cow
paradise
blue



Male-Female



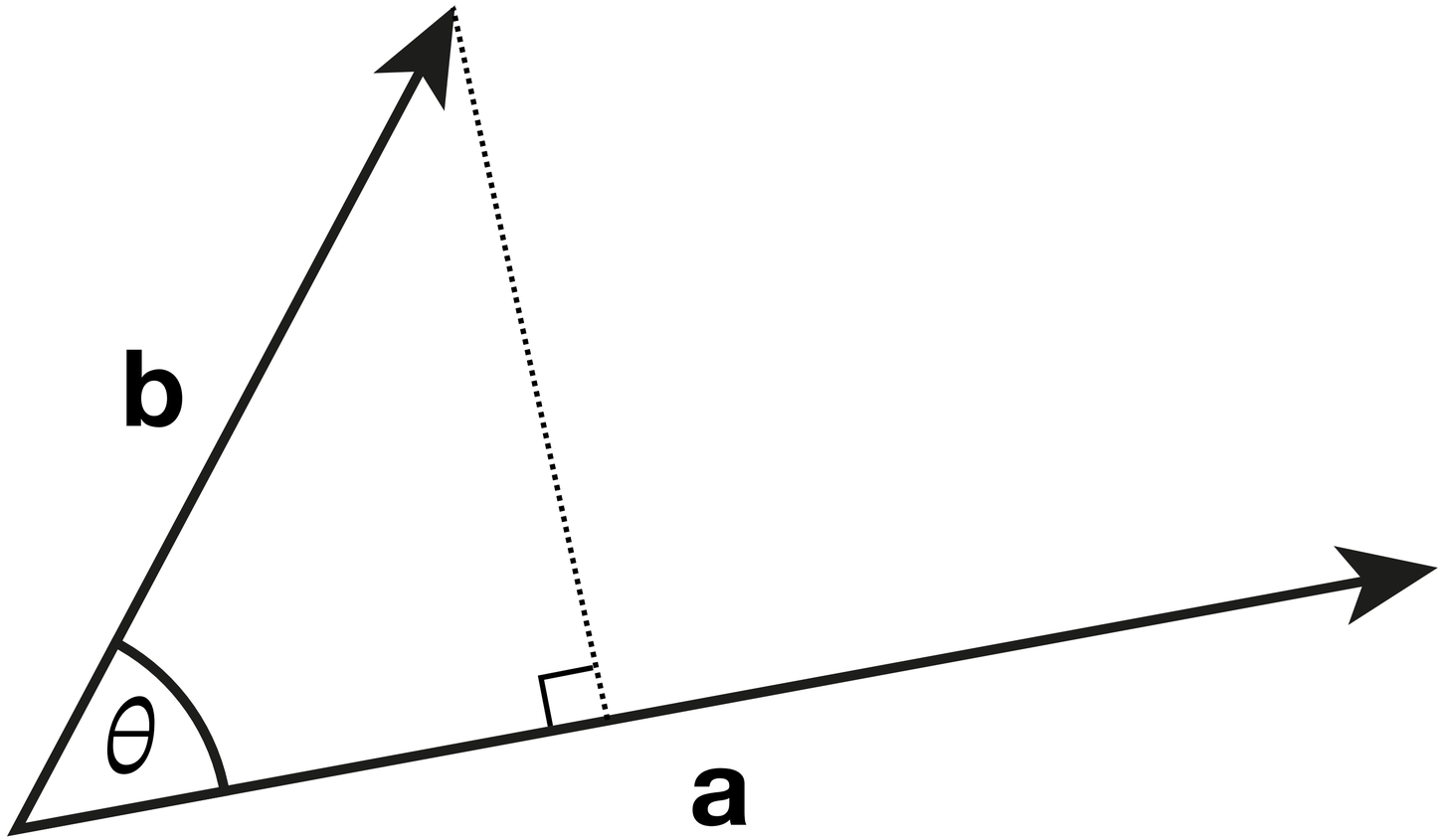
Verb Tense

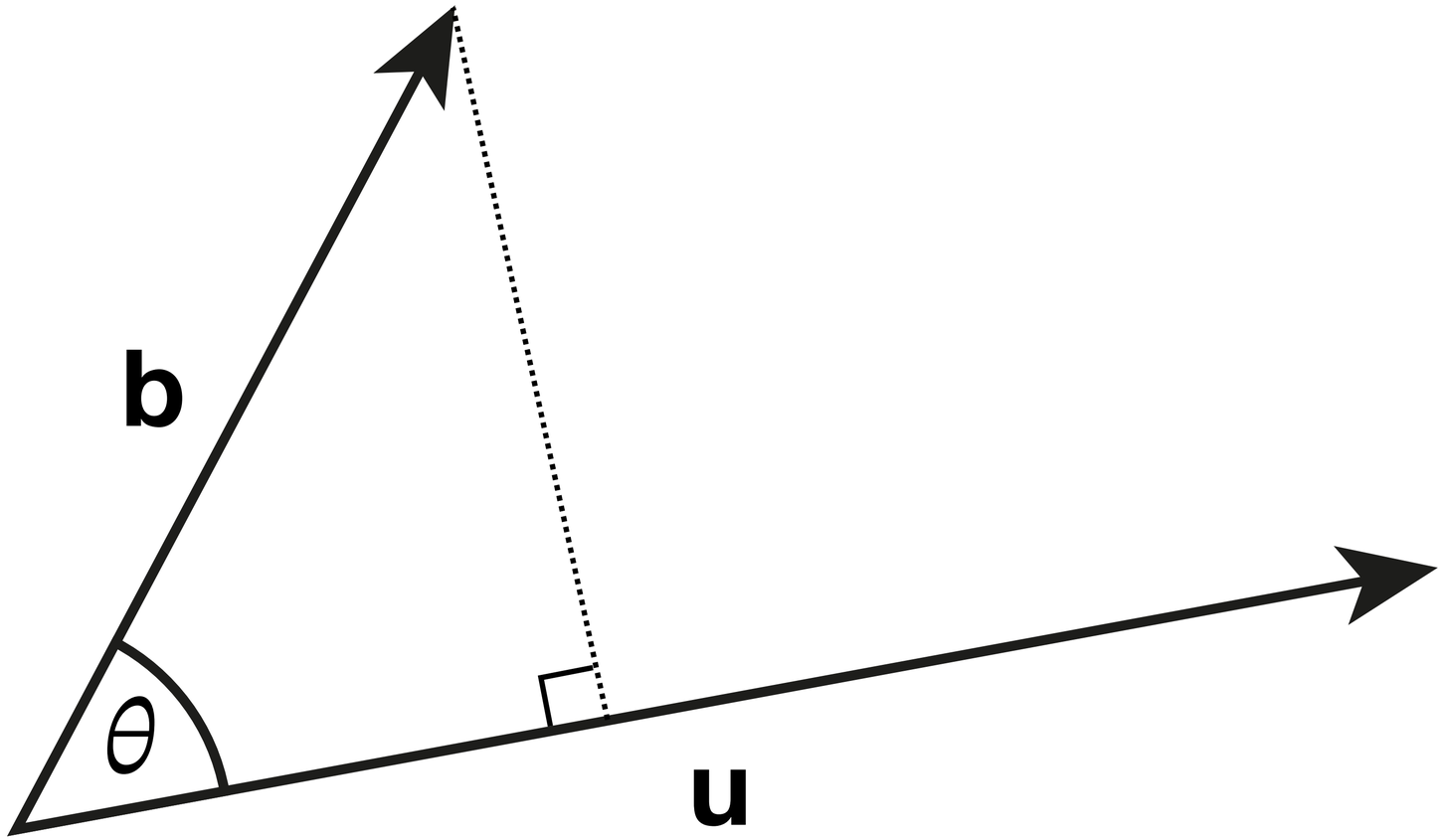


Country-Capital

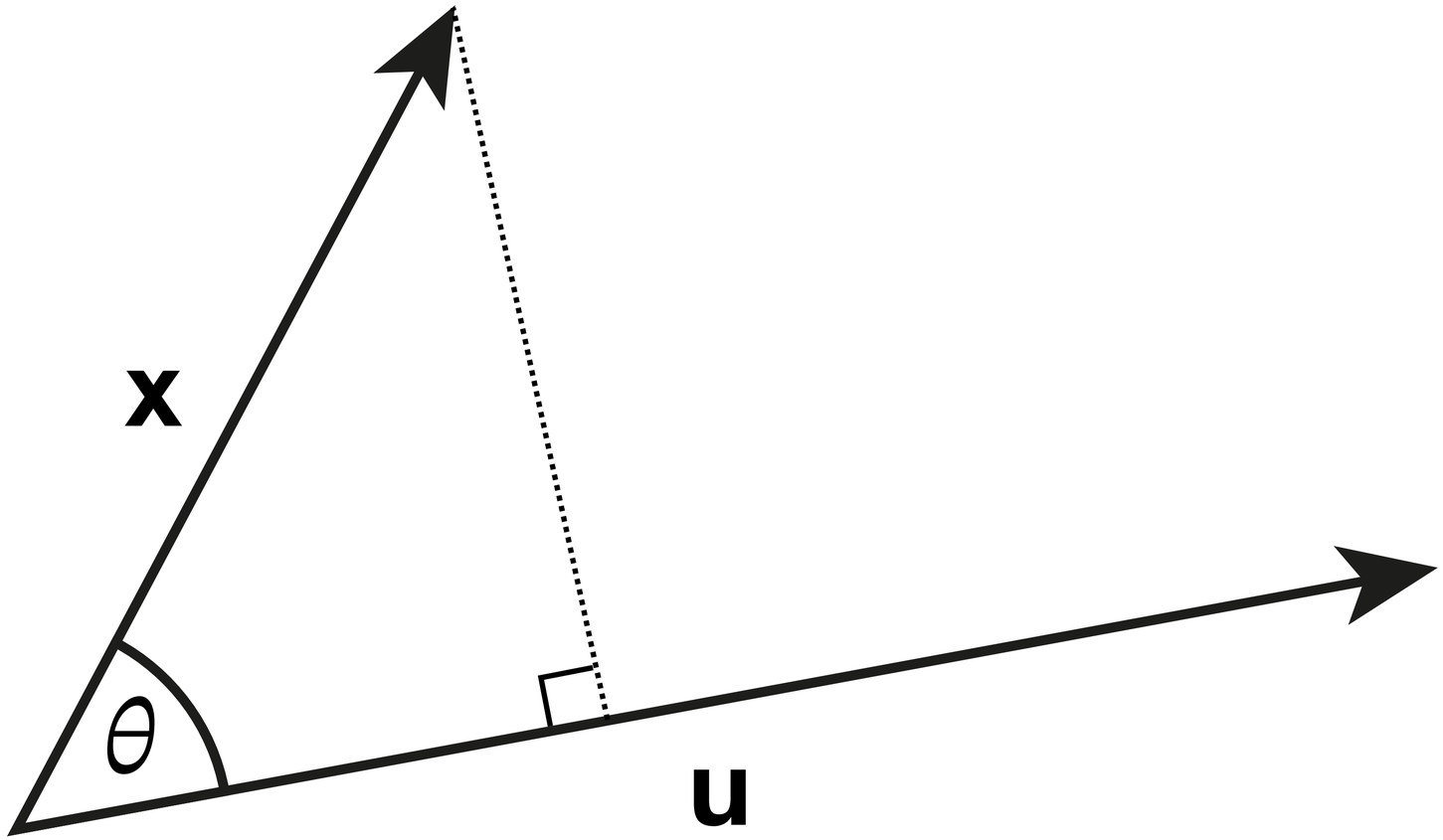
PCA

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos \theta$$


b
θ
a



$$|\mathbf{u}| = 1$$



$$|\mathbf{u}| = 1$$

$$\mathbf{x} \cdot \mathbf{u} = |\mathbf{x}| \cos \theta$$

$$\mathbf{x} \cdot \mathbf{u}_1 = |\mathbf{x}| \cos \theta_1 = \tilde{x}_1$$

$$\mathbf{x} \cdot \mathbf{u}_1 = |\mathbf{x}| \cos\theta_1 = \tilde{x}_1$$

$$\mathbf{x} \cdot \mathbf{u}_2 = |\mathbf{x}| \cos\theta_2 = \tilde{x}_2$$

$$\mathbf{x} \cdot \mathbf{u}_1 = |\mathbf{x}| \cos\theta_1 = \tilde{x}_1$$

$$\mathbf{x} \cdot \mathbf{u}_2 = |\mathbf{x}| \cos\theta_2 = \tilde{x}_2$$

$$\vdots$$

$$\mathbf{x} \cdot \mathbf{u}_p = |\mathbf{x}| \cos\theta_p = \tilde{x}_p$$

$$\mathbf{x}^\top \mathbf{U} = \tilde{\mathbf{x}}$$

$$\mathbf{XU} = \tilde{\mathbf{X}}$$

$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

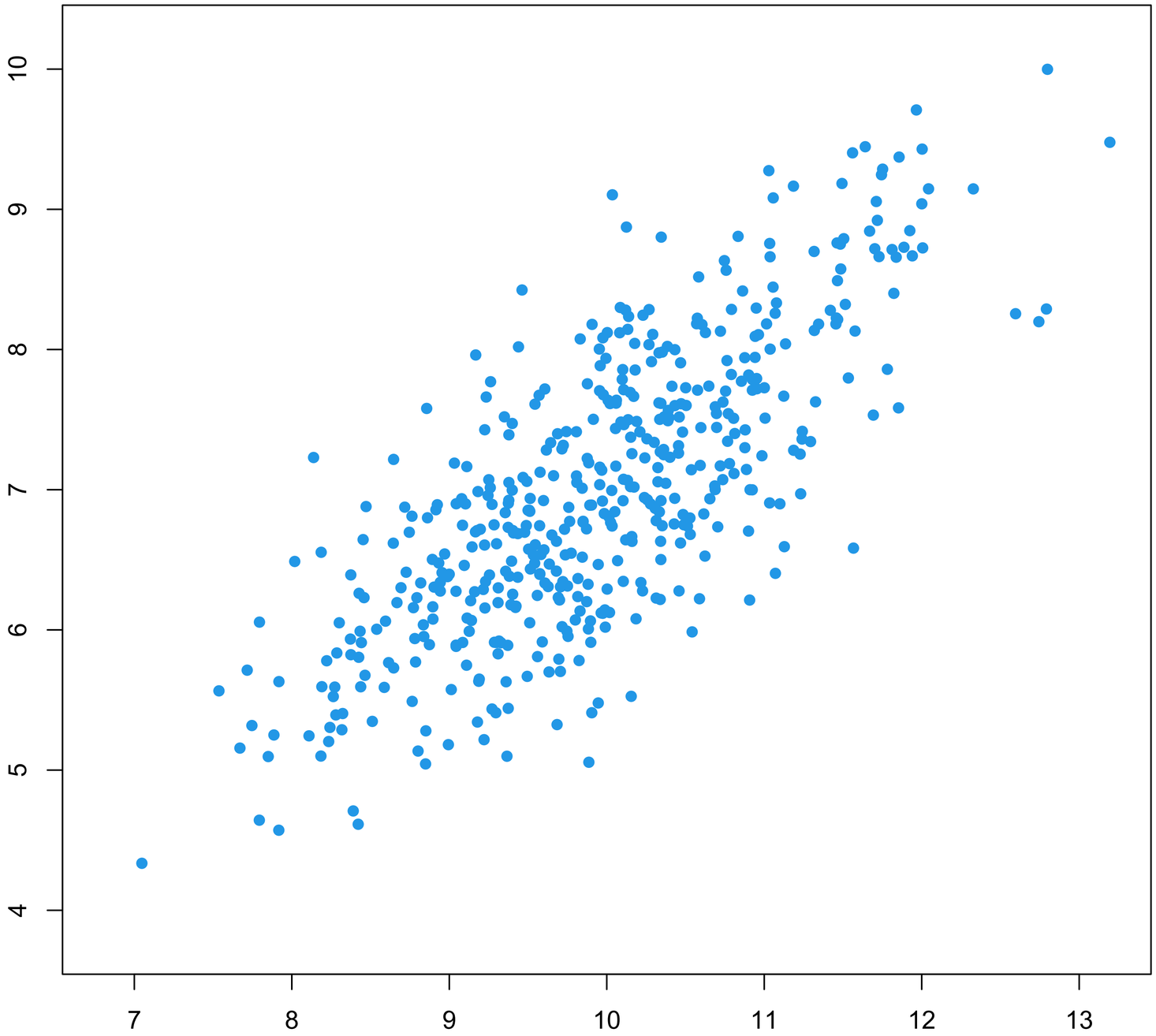
$$\mathbf{U} \in \mathbb{R}^{d \times p}$$

$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

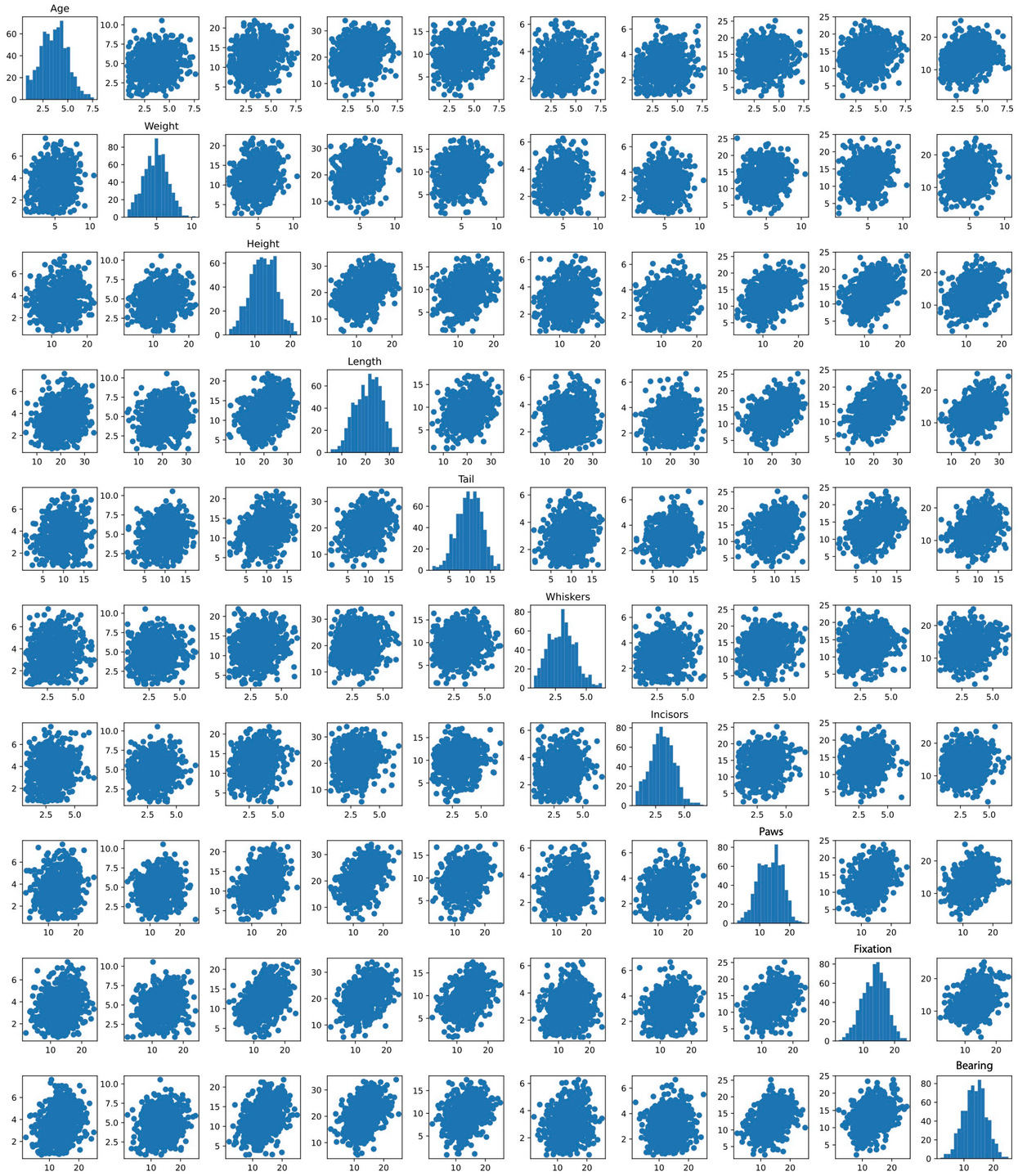
$$\mathbf{U} \in \mathbb{R}^{d \times p}$$

$$\tilde{\mathbf{X}} = \mathbf{X}\mathbf{U} \in \mathbb{R}^{n \times p}$$

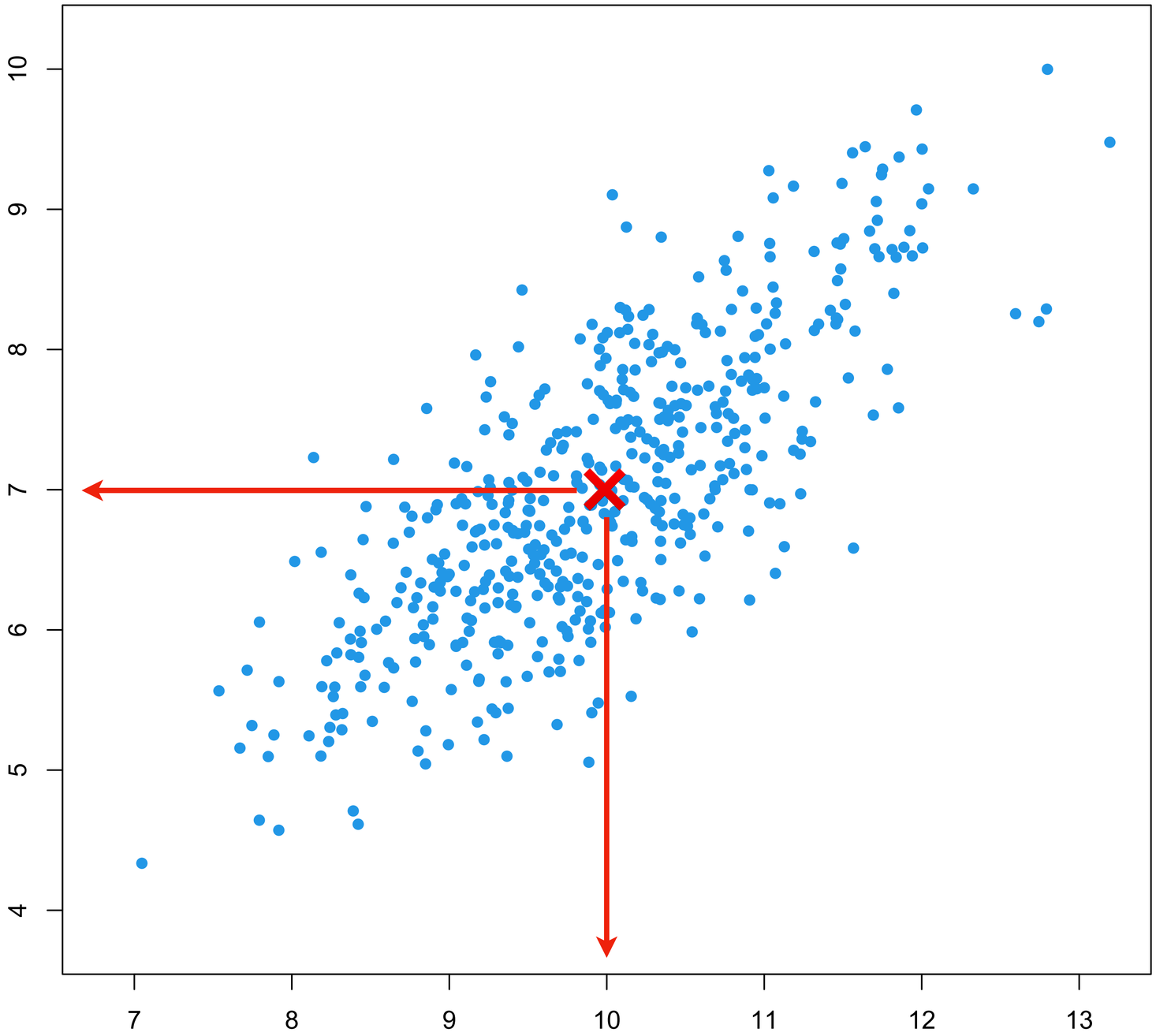
$$\tilde{\mathbf{X}} = \mathbf{X}\mathbf{U}$$

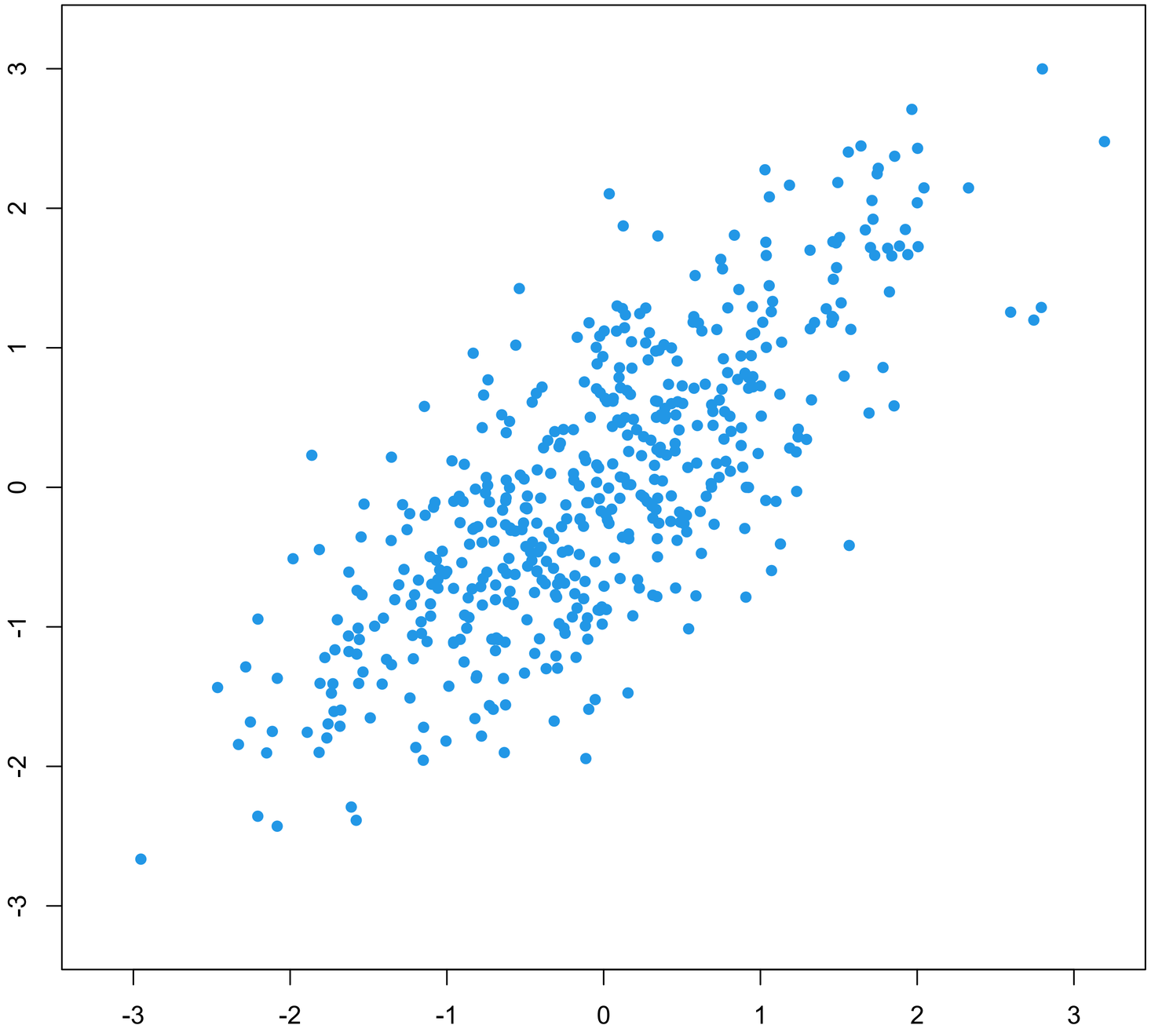

10
9
8
7
6
5
4
7
8
9
10
11
12
13

$$\Sigma = \begin{bmatrix} \mathrm{var}(\mathbf{x}_{|1}) & \mathrm{covar}(\mathbf{x}_{|1}, \mathbf{x}_{|2}) & \cdots & \mathrm{covar}(\mathbf{x}_{|1}, \mathbf{x}_{|d}) \\ \mathrm{covar}(\mathbf{x}_{|2}, \mathbf{x}_{|1}) & \mathrm{var}(\mathbf{x}_{|2}) & \cdots & \mathrm{covar}(\mathbf{x}_{|2}, \mathbf{x}_{|d}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{covar}(\mathbf{x}_{|d}, \mathbf{x}_{|1}) & \mathrm{covar}(\mathbf{x}_{|d}, \mathbf{x}_{|2}) & \cdots & \mathrm{var}(\mathbf{x}_{|d}) \end{bmatrix}$$

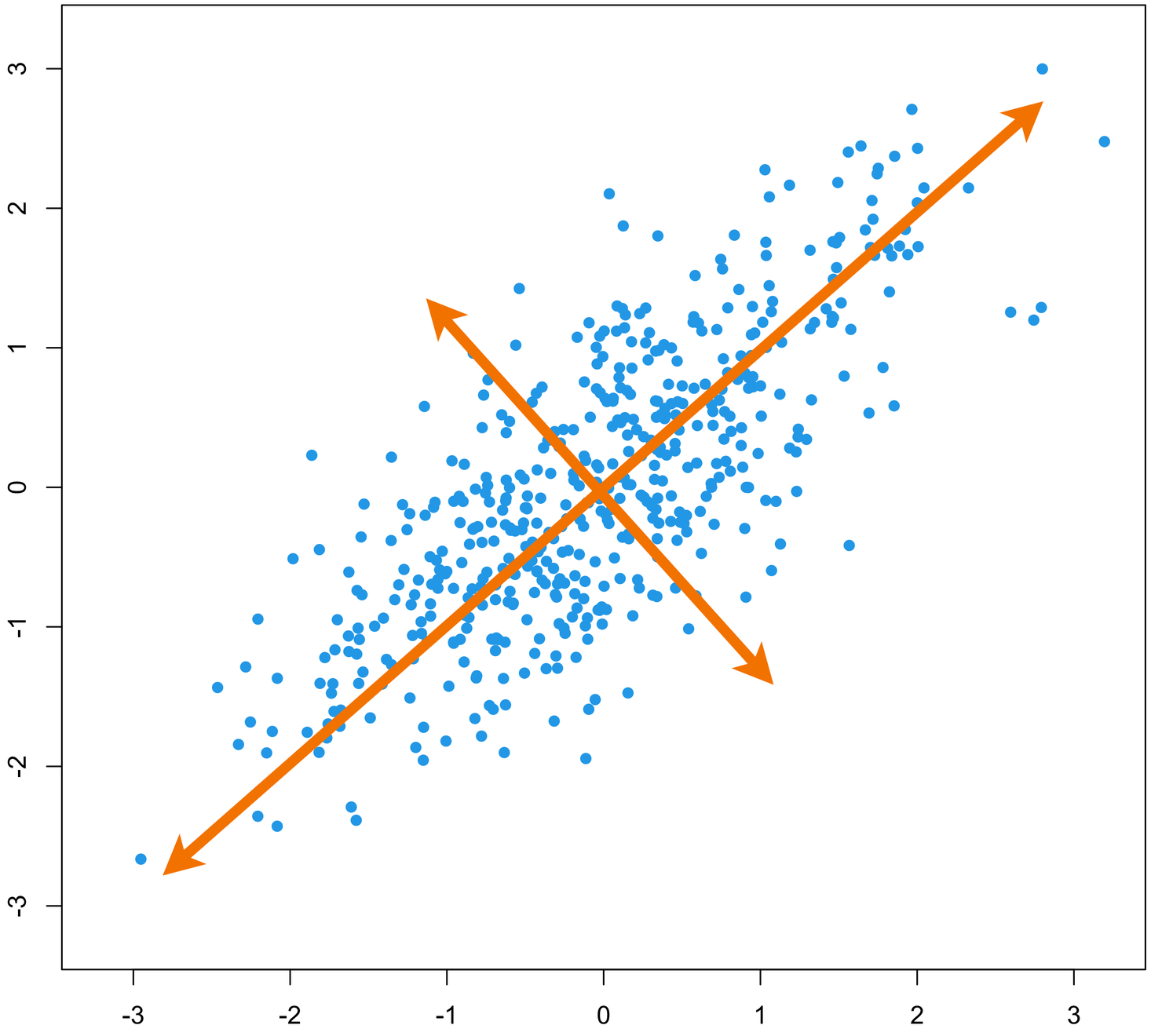

Age
Weight
Height
Length
Tail
Whiskers
Incisors
Paws
Fixation
Bearing

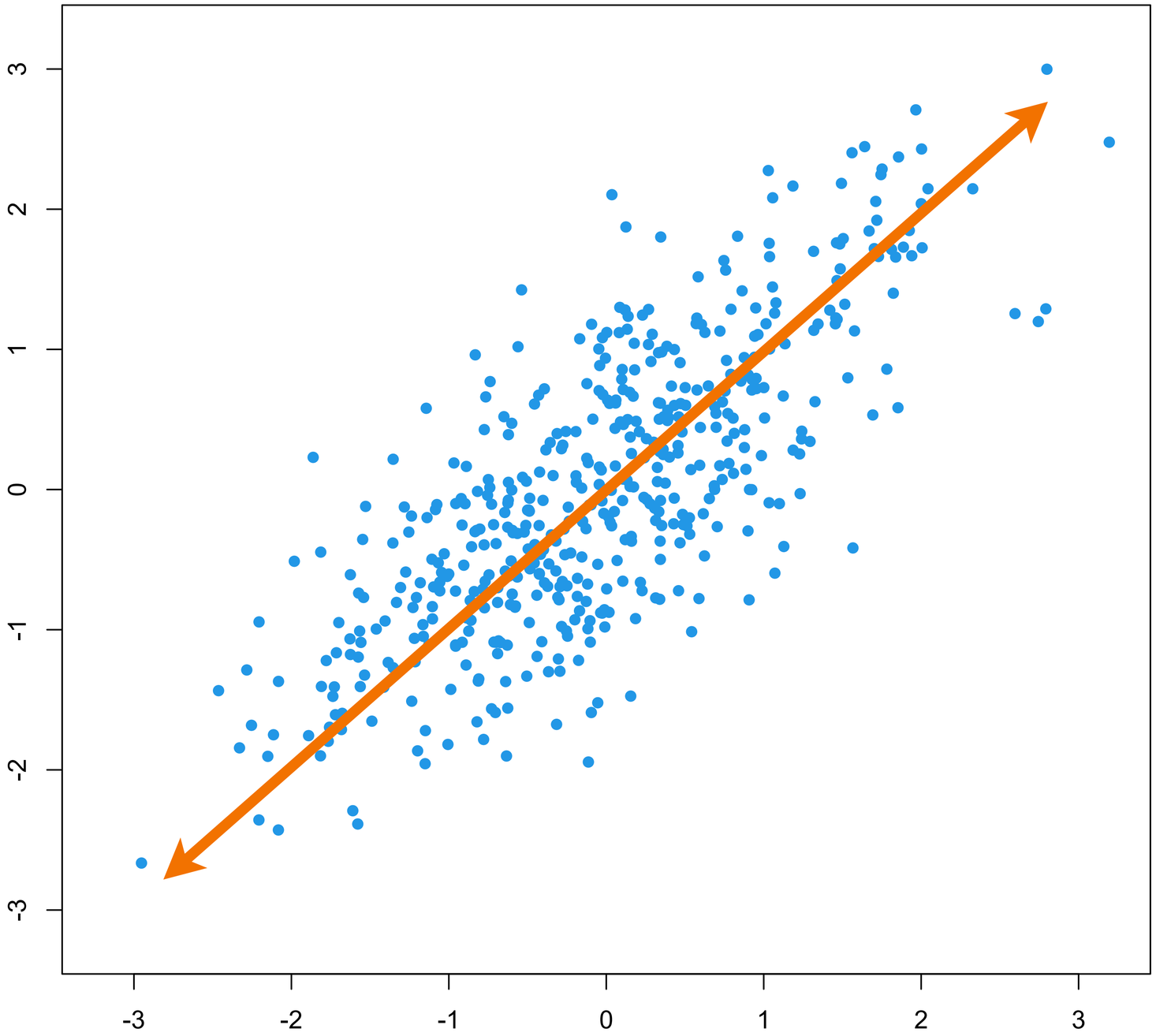
$$\Sigma = \frac{1}{n}\mathbf{X}^{\top}\mathbf{X}$$


10
9
8
7
6
5
4
7
8
9
10
11
12
13


3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

$$\mathbf{M}\mathbf{v} = \lambda\mathbf{v}$$


3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

$$\tilde{x}_i = \mathbf{u} \cdot \mathbf{x}_i$$

$$\tilde{\mathbf{x}}_i = \tilde{x}_i \mathbf{u} = \left(\mathbf{u} \cdot \mathbf{x}_i\right)\mathbf{u}$$

$$\mathrm{var}(\tilde{\mathbf{x}}) = \frac{1}{n} \sum_i \tilde{x}_i^2$$

$$
\begin{aligned}
\mathrm{var}(\tilde{\mathbf{X}}) &= \frac{1}{n}\sum_i \tilde{x}_i^2 \\
&= \frac{1}{n}\sum_i (\mathbf{u}\cdot\mathbf{x}_i)^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(\tilde{\mathbf{X}}) &= \frac{1}{n}\sum_i \tilde{x}_i^2 \\
&= \frac{1}{n}\sum_i (\mathbf{u}\cdot\mathbf{x}_i)^2 \\
&= \frac{1}{n}\mathbf{u}^\top \mathbf{X}^\top \mathbf{X}\mathbf{u}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(\tilde{\mathbf{X}}) &= \frac{1}{n}\sum_i \tilde{x}_i^2 \\
&= \frac{1}{n}\sum_i (\mathbf{u}\cdot\mathbf{x}_i)^2 \\
&= \frac{1}{n}\mathbf{u}^\top \mathbf{X}^\top \mathbf{X}\mathbf{u} \\
&= \mathbf{u}^\top \boldsymbol{\Sigma}\mathbf{u}
\end{aligned}
$$

$$\underset{\mathbf{u}}{\text{maximise}} \quad \mathbf{u}^\top \boldsymbol{\Sigma} \mathbf{u}$$

$$\underset{\mathbf{u}}{\text{maximise}} \;\; \mathbf{u}^\top \boldsymbol{\Sigma} \mathbf{u}$$

$$\text{subject to: } \|\mathbf{u}\| = 1$$

$$\mathbf{u}^\top \boldsymbol{\Sigma} \mathbf{u} + \lambda(1 - \mathbf{u}^\top \mathbf{u})$$

$$\mathbf{u}^\top \boldsymbol{\Sigma} \mathbf{u} + \lambda(1 - \mathbf{u}^\top \mathbf{u})$$

$$2\boldsymbol{\Sigma}\mathbf{u} - 2\lambda\mathbf{u}$$

$$\mathbf{u}^\top \boldsymbol{\Sigma} \mathbf{u} + \lambda(1 - \mathbf{u}^\top \mathbf{u})$$

$$2\boldsymbol{\Sigma}\mathbf{u} - 2\lambda\mathbf{u} = 0$$

$$\mathbf{u}^\top \boldsymbol{\Sigma} \mathbf{u} + \lambda(1 - \mathbf{u}^\top \mathbf{u})$$

$$2\boldsymbol{\Sigma}\mathbf{u} - 2\lambda\mathbf{u} = 0$$

$$\boldsymbol{\Sigma}\mathbf{u} = \lambda\mathbf{u}$$

$$\mathbf{u}^\top \boldsymbol{\Sigma} \mathbf{u} = \lambda \mathbf{u}^\top \mathbf{u}$$

$$\mathbf{u}^\top \boldsymbol{\Sigma} \mathbf{u} = \lambda \mathbf{u}^\top \mathbf{u} = \lambda$$

$\mathbf{U} \in \mathbb{R}^{d \times d}$ cols eigenvectors of $\mathbf{X}^\top \mathbf{X}$

$$\mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{U}^\top$$

$\mathbf{X} \in \mathbb{R}^{n \times d}$ original data

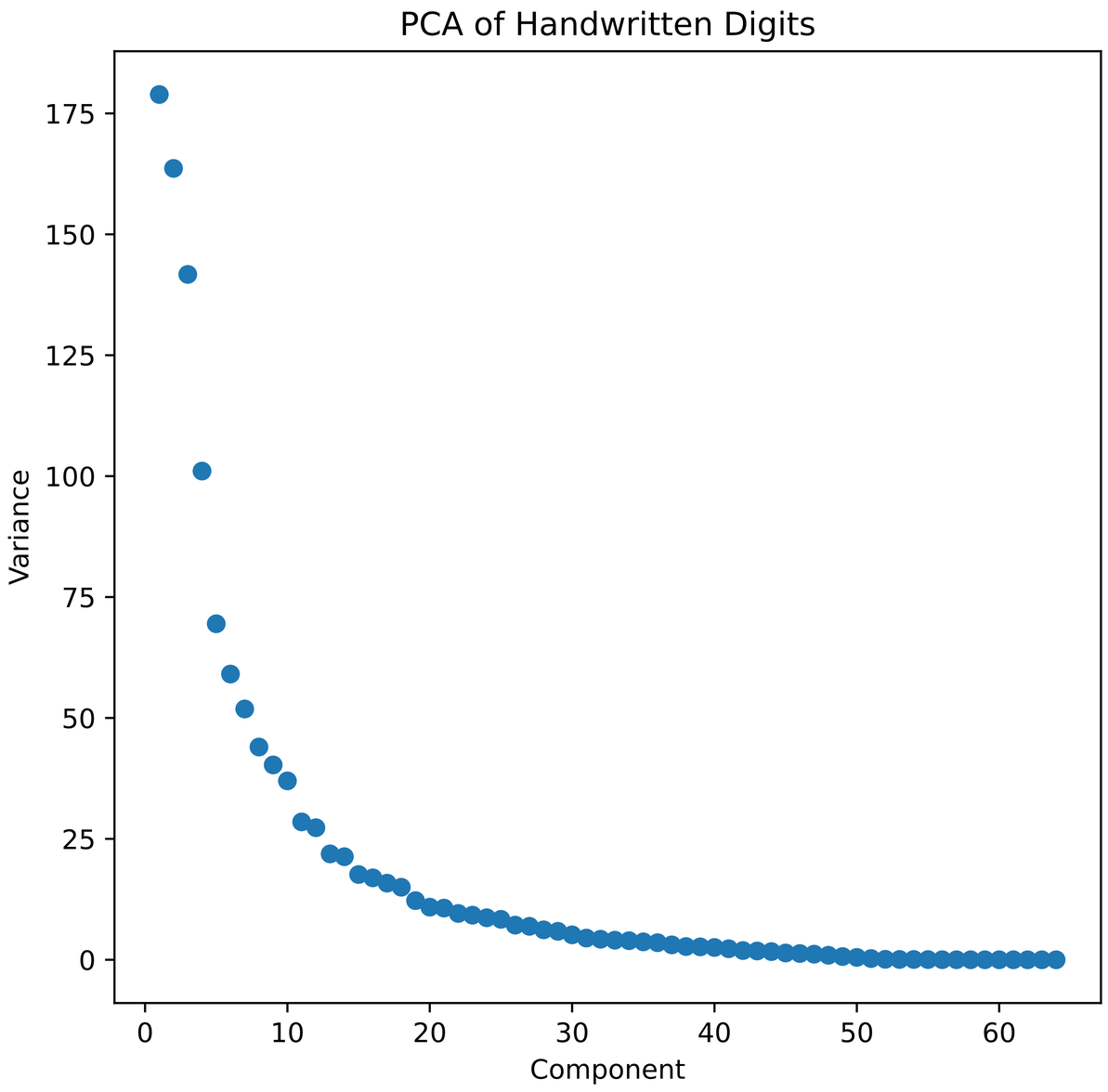
$\mathbf{V} \in \mathbb{R}^{n \times n}$ cols are eigenvectors of $\mathbf{X}\mathbf{X}^\top$

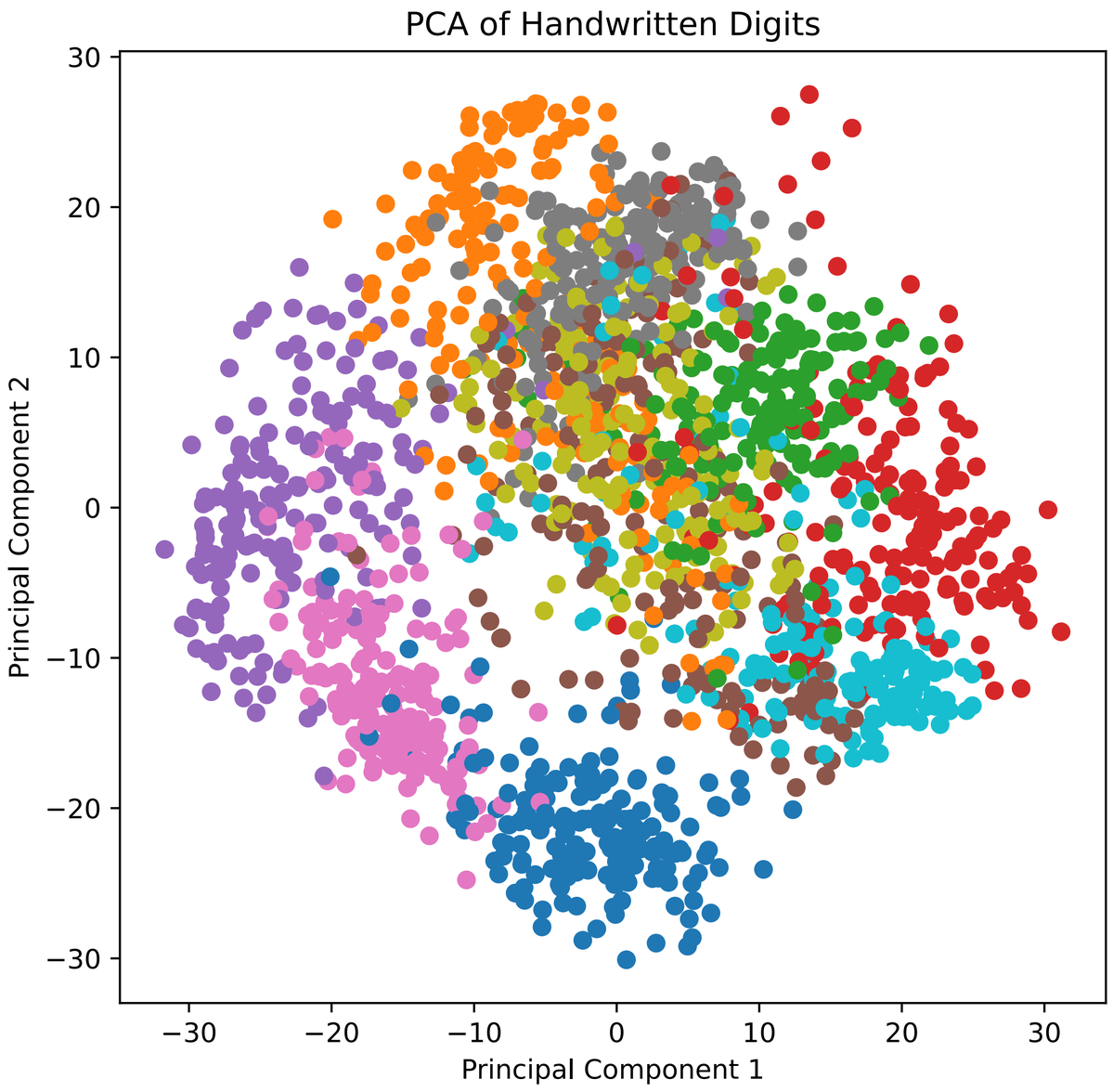
$\mathbf{D} \in \mathbb{R}^{n \times d}$ $\sqrt{\mathrm{diag}(\mathrm{eig}(\mathbf{X}\mathbf{X}^\top))}$

$\mathbf{U} \in \mathbb{R}^{d \times d}$ cols eigenvectors of $\mathbf{X}^\top\mathbf{X}$

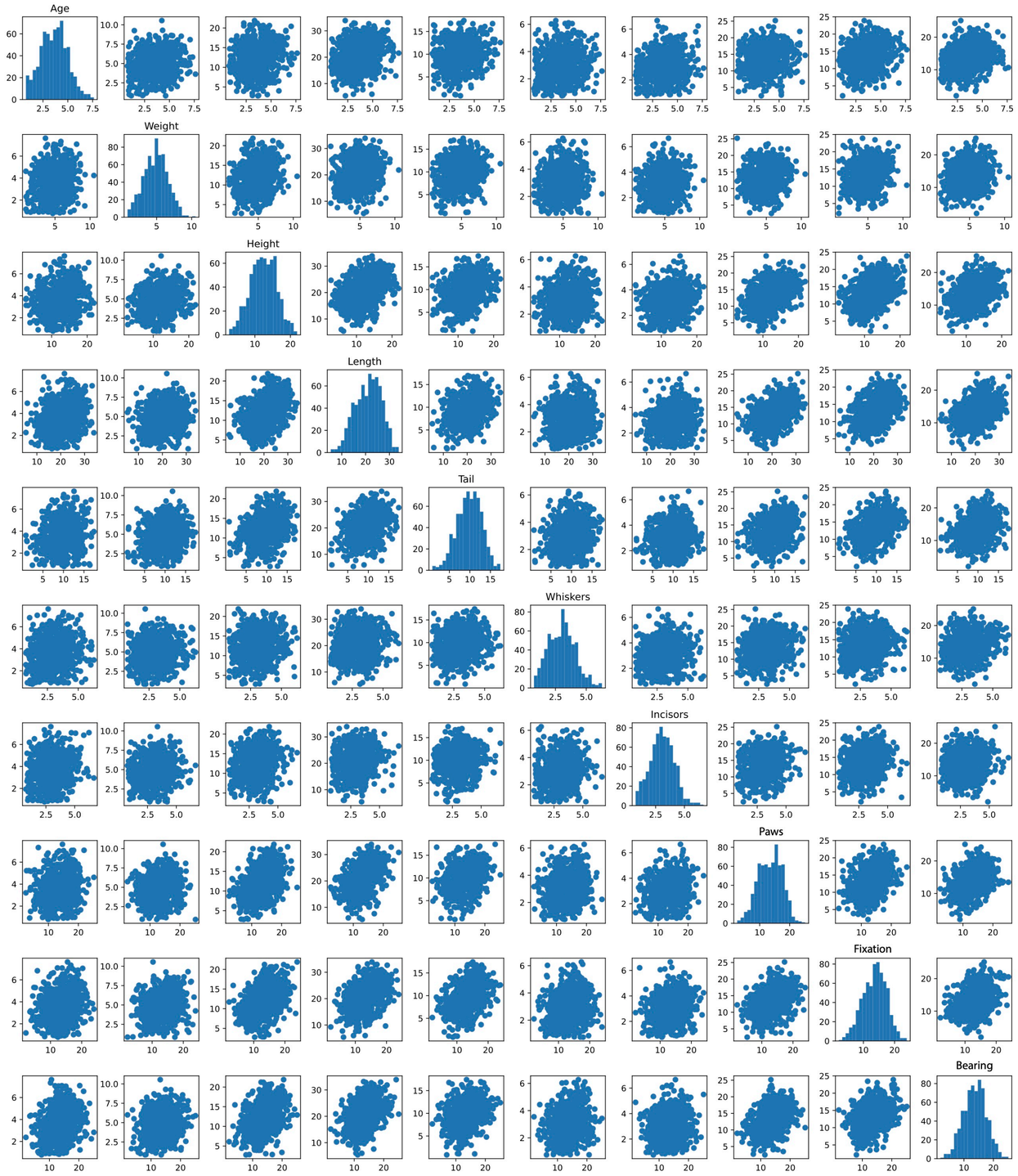
$$\mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{U}^\top$$

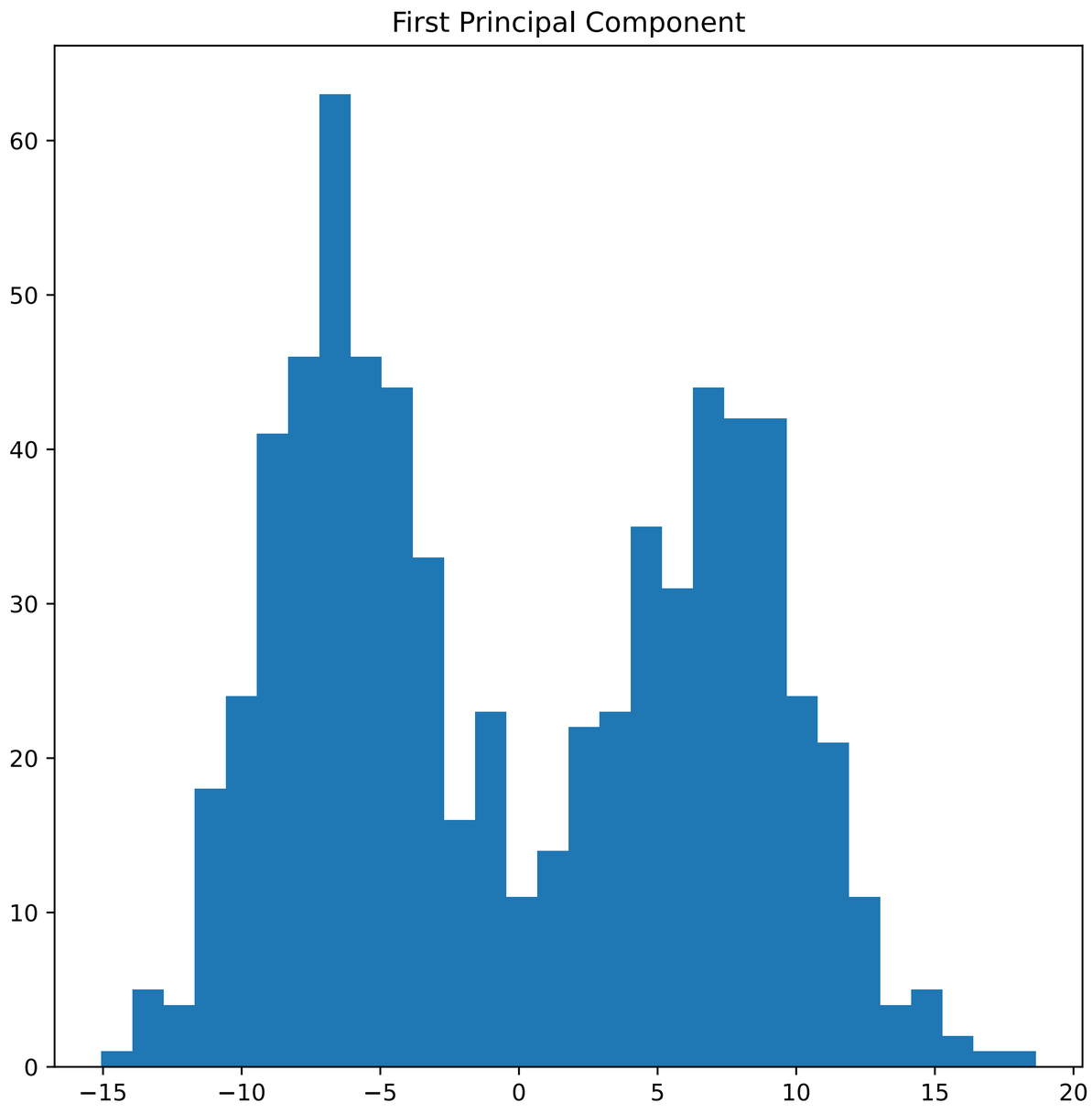
$$\begin{aligned}\tilde{\mathbf{X}} &= \mathbf{X}\mathbf{U} \\ &= \mathbf{V}\mathbf{D}\mathbf{U}^\top\mathbf{U} \\ &= \mathbf{V}\mathbf{D}\end{aligned}$$

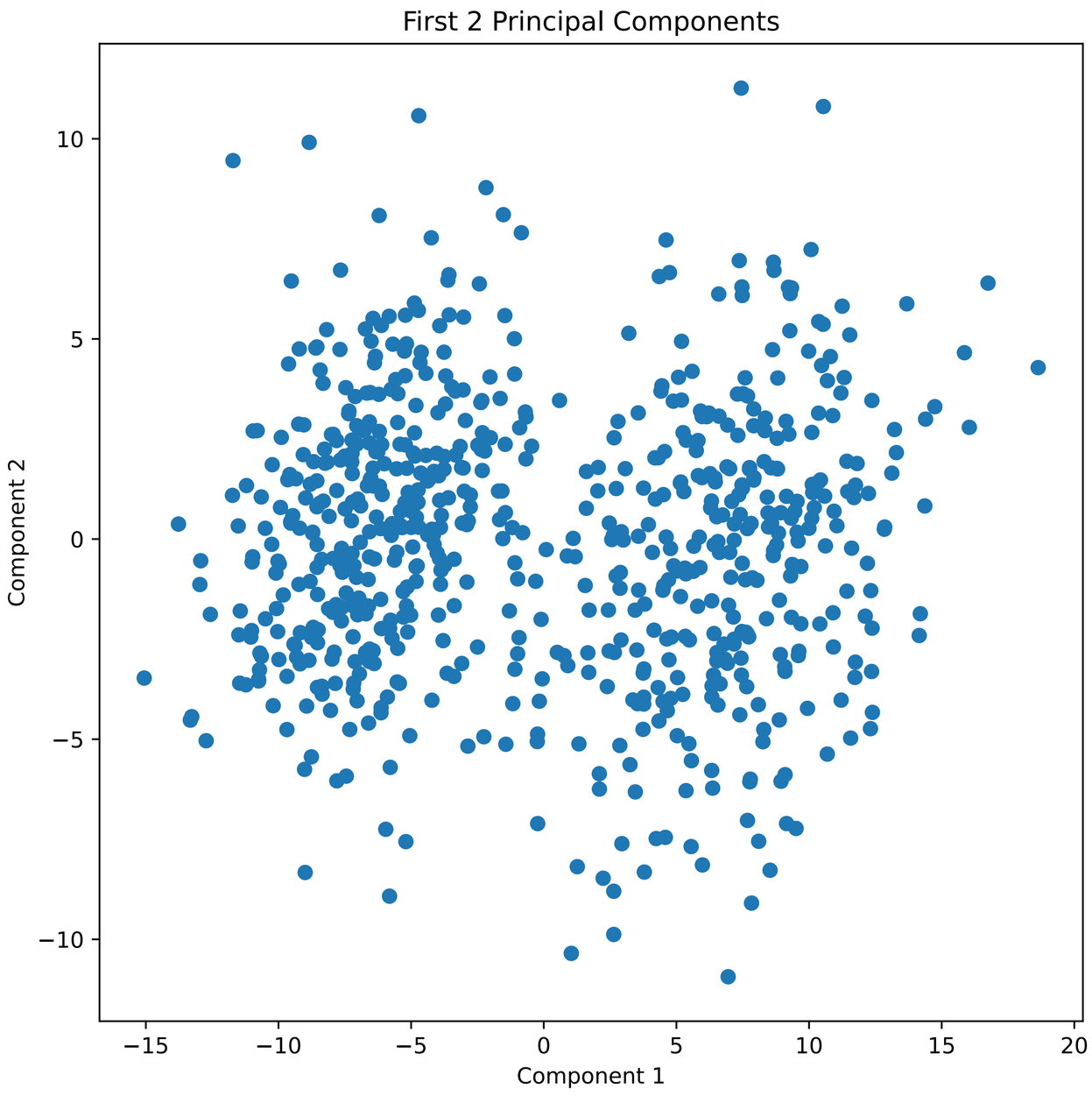

PCA of Handwritten Digits
Variance
Component
0
25
50
75
100
125
150
175
0
10
20
30
40
50
60


PCA of Handwritten Digits
Principal Component 2
Principal Component 1
30
20
10
0
−10
−20
−30
−30
−20
−10
0
10
20
30

Original Images
Mean Only
1 Component
2 Components
3 Components
4 Components
5 Components
6 Components
7 Components
8 Components
9 Components
10 Components
11 Components
12 Components
13 Components
14 Components
15 Components
16 Components
17 Components
18 Components
19 Components
20 Components
21 Components
22 Components


Age
Weight
Height
Length
Tail
Whiskers
Incisors
Paws
Fixation
Bearing


First Principal Component
60
50
40
30
20
10
0
−15
−10
−5
0
5
10
15
20


First 2 Principal Components
Component 2
Component 1
10
5
0
−5
−10
−15
−10
−5
0
5
10
15
20

clustering

$$\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_k\} \quad \text{cluster centroids} \in \mathbb{R}^d$$

$$\mathbf{a} = \{a_1, a_2, ..., a_n\} \quad \text{cluster assignments} \ \in \{1, 2, ..., k\}$$

$$\hat{\mathbf{C}}, \hat{\mathbf{a}} = \underset{\mathbf{C},\mathbf{a}}{\operatorname{argmin}} \sum_{i}^{n} \|\mathbf{x}_i - \mathbf{c}_{a_i}\|^2$$

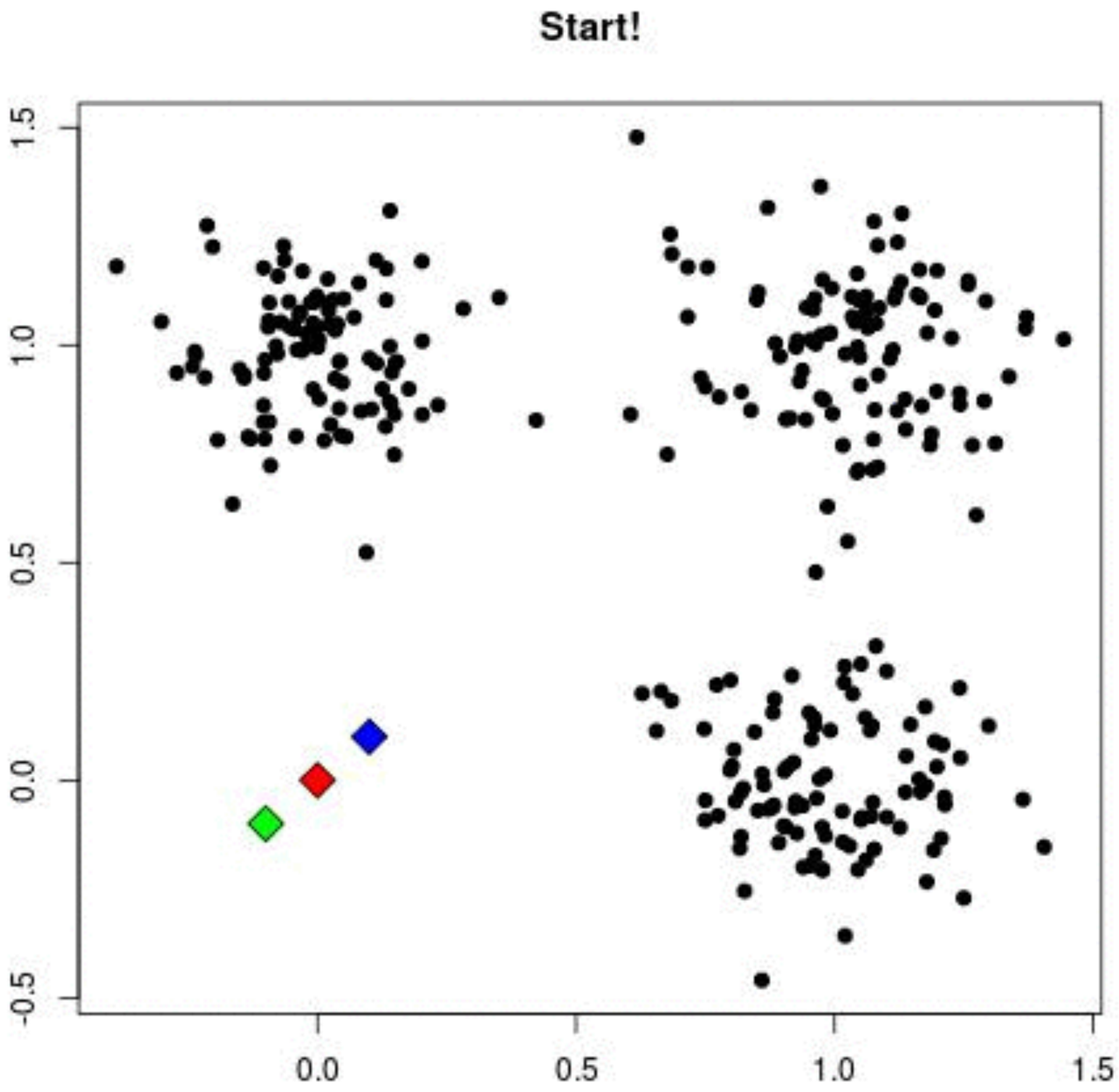
Initialise **c** somehow

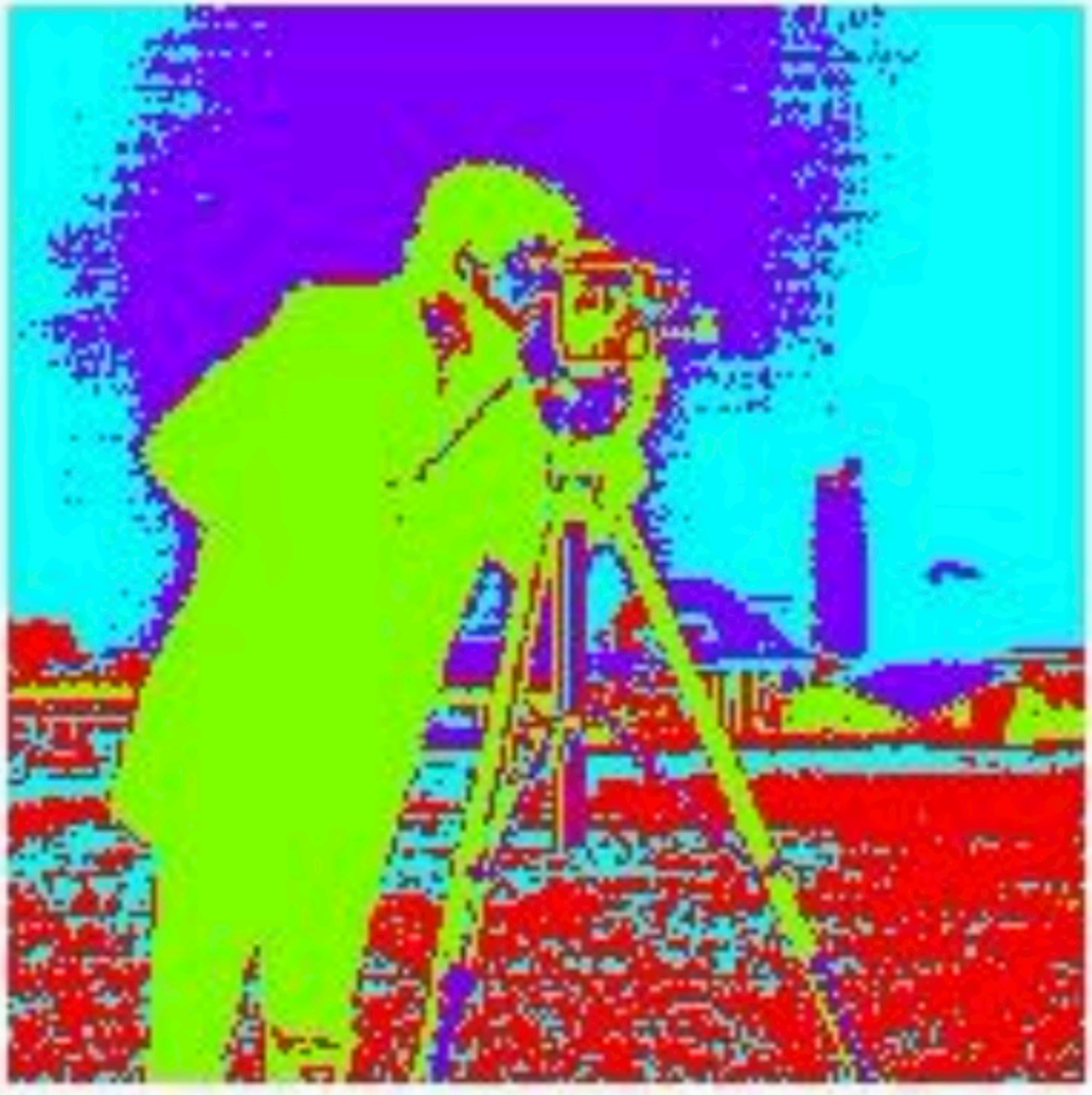
Repeat

  Assign each **x** to its nearest **c**

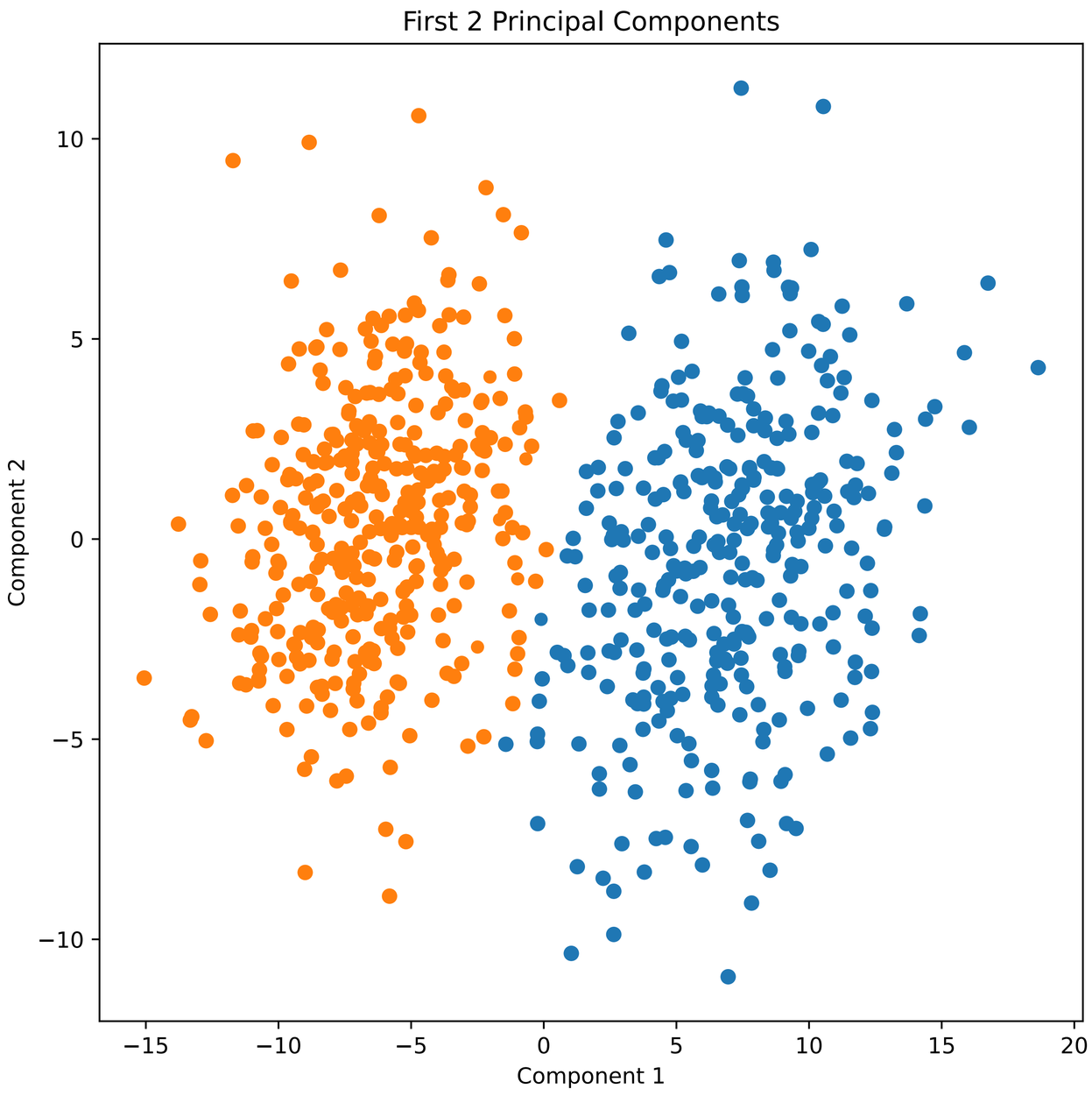
  Update each **c** to be the mean of all **x** assigned to it

Until nothing changes


Start!
1.5
1.0
0.5
0.0
-0.5
0.0
0.5
1.0
1.5



Agglomerative
Divisive


First 2 Principal Components
Component 2
Component 1
10
5
0
−5
−10
−15
−10
−5
0
5
10
15
20

Questions?

Next: Expectation Maximisation