

Week 7: Clustering and Dimensionality Reduction

Matthew Caldwell

COMP0088 Introduction to Machine Learning • UCL Computer Science • Autumn 2023

Admin

- Pulse surveys feedback

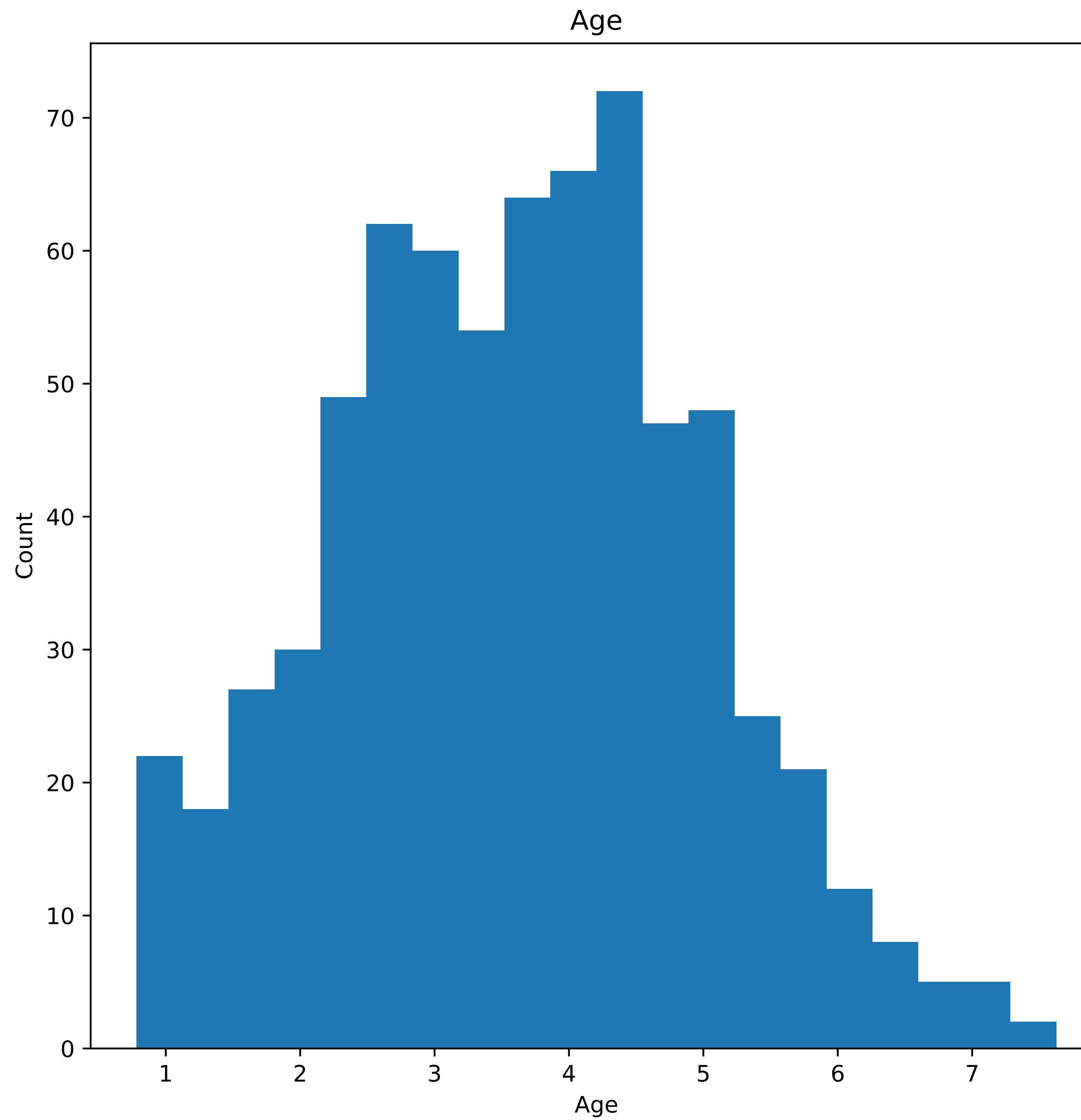
Week 7 Recap

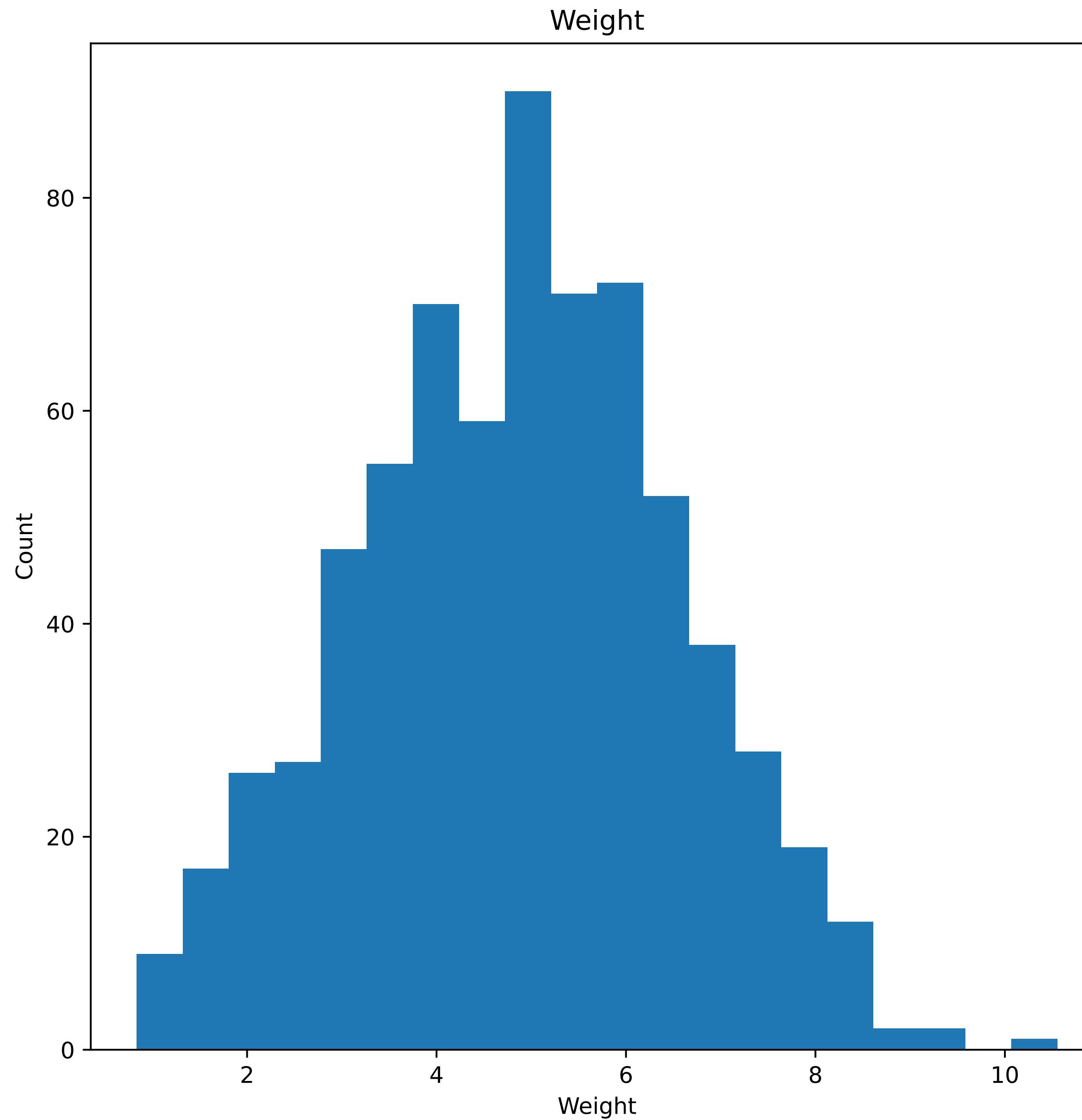
Transylvania 6-5000

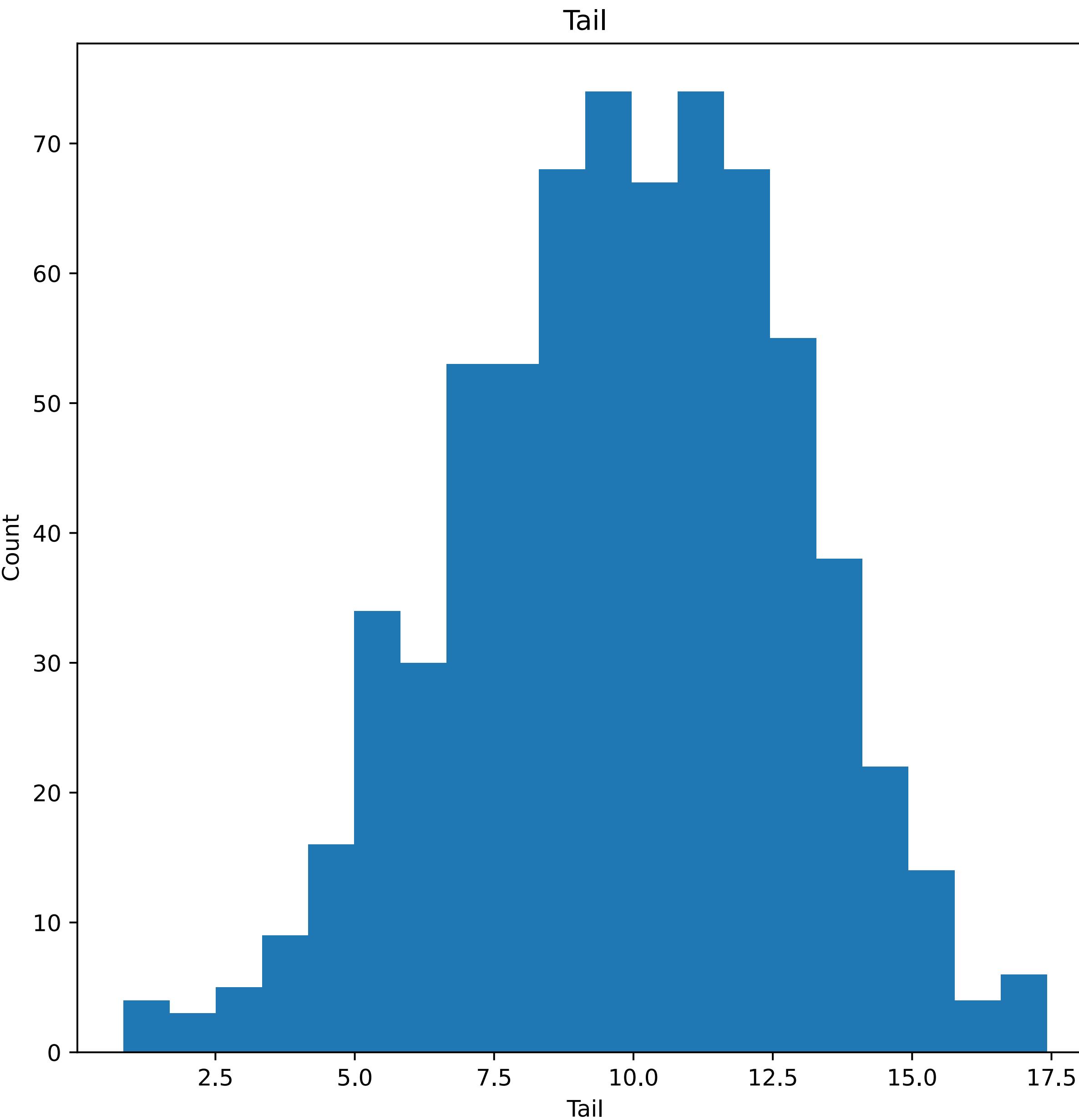


groundhogs

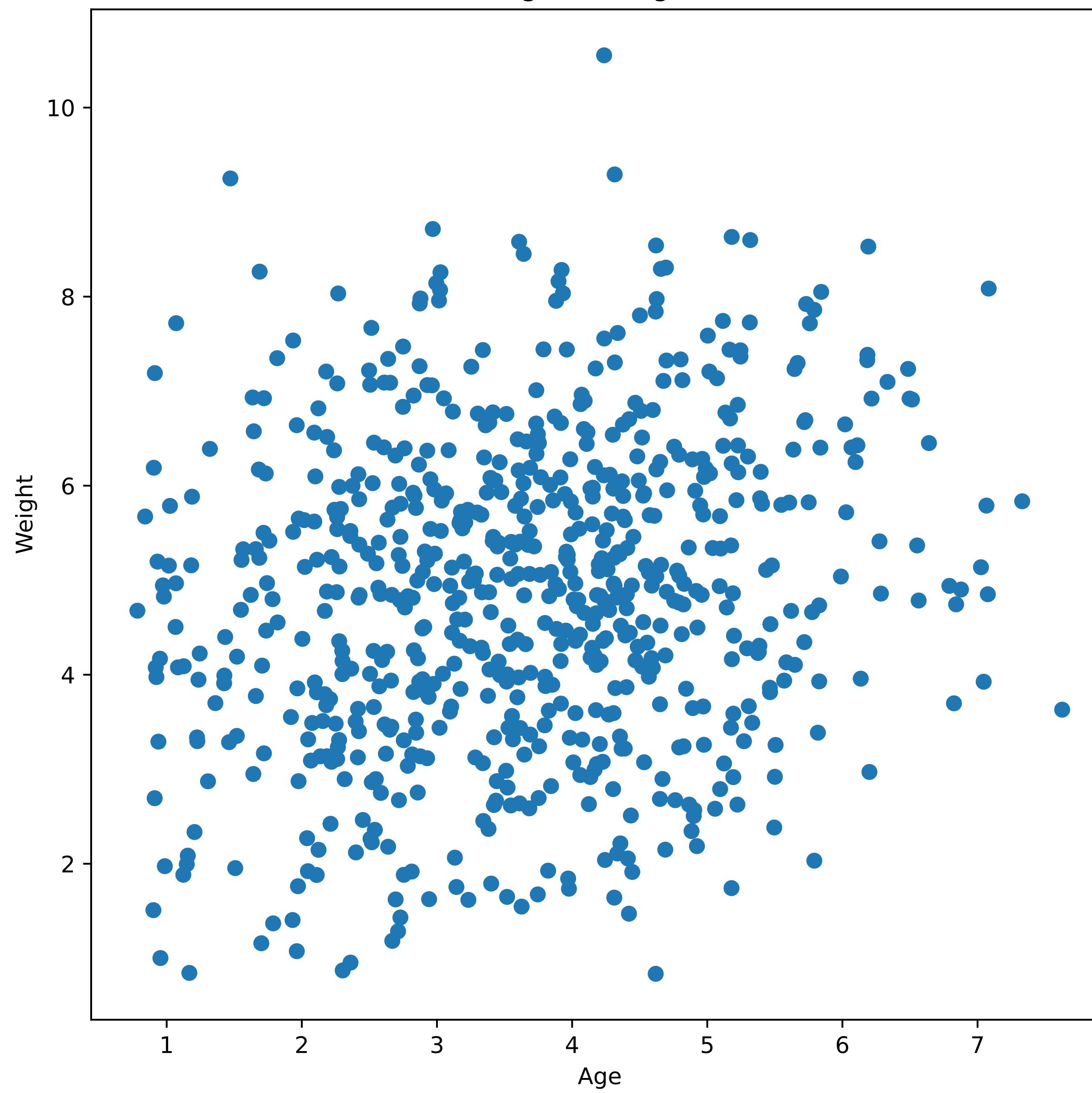
	Age	Weight	Height	Length	Tail	Whiskers	Incisors	E
0	2.2	3.8	14.3	14.7	5.5	4.5	1.4	
1	4.9	2.6	11.8	13.3	8.3	0.9	2.2	
2	2.6	5.2	12.6	15.9	9.9	1.9	3.6	
3	7.5	4.3	7.0	15.2	9.3	0.5	1.8	
4	2.2	3.1	11.4	15.0	12.0	1.8	2.6	
5	5.2	6.1	13.2	15.7	5.6	3.3	3.3	
6	3.5	3.1	7.7	20.2	7.4	0.7	2.9	
7	3.7	1.7	2.8	15.3	12.7	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	
	5.9	4.7	6.5	16.4	--		4.6	

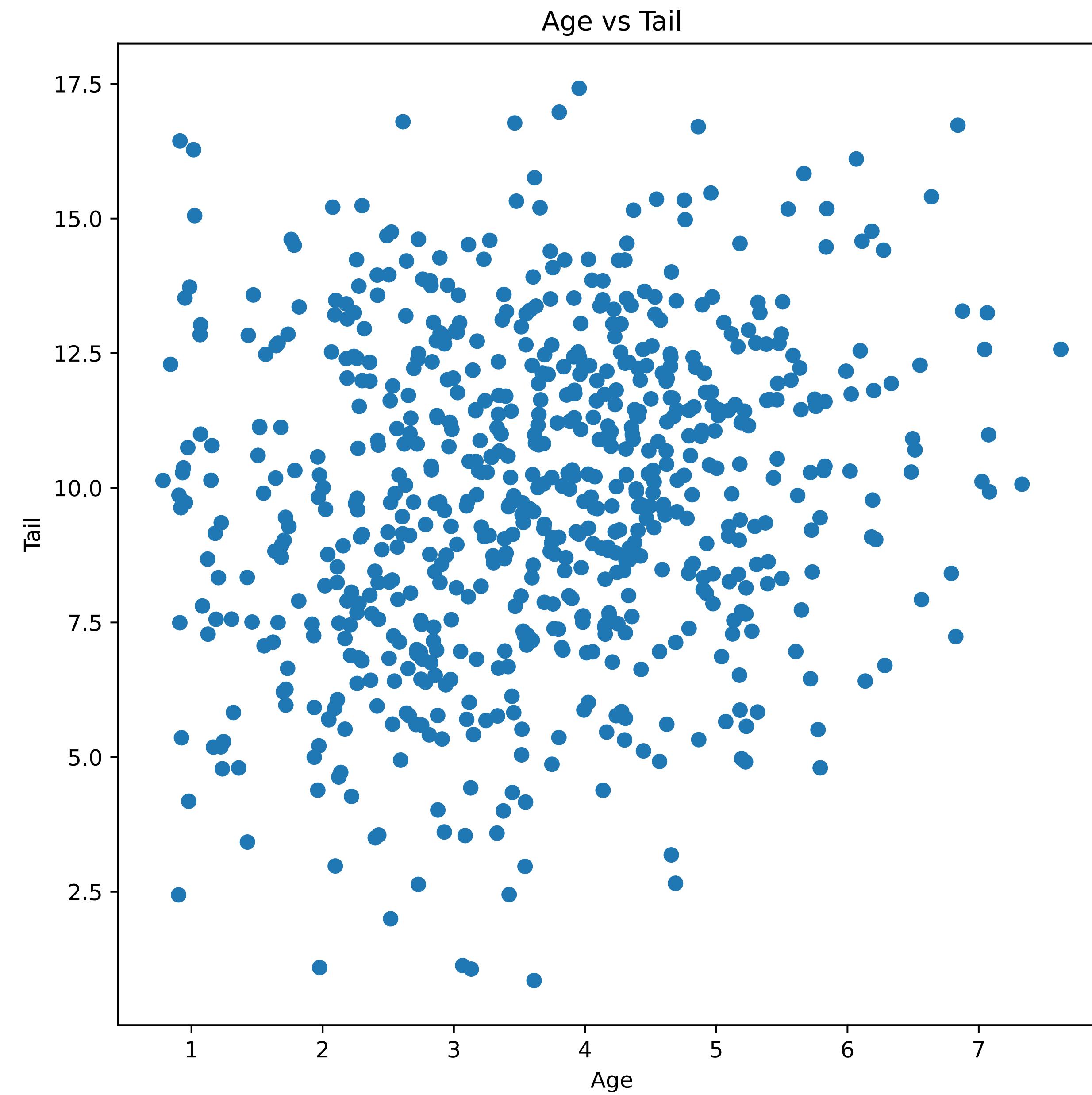




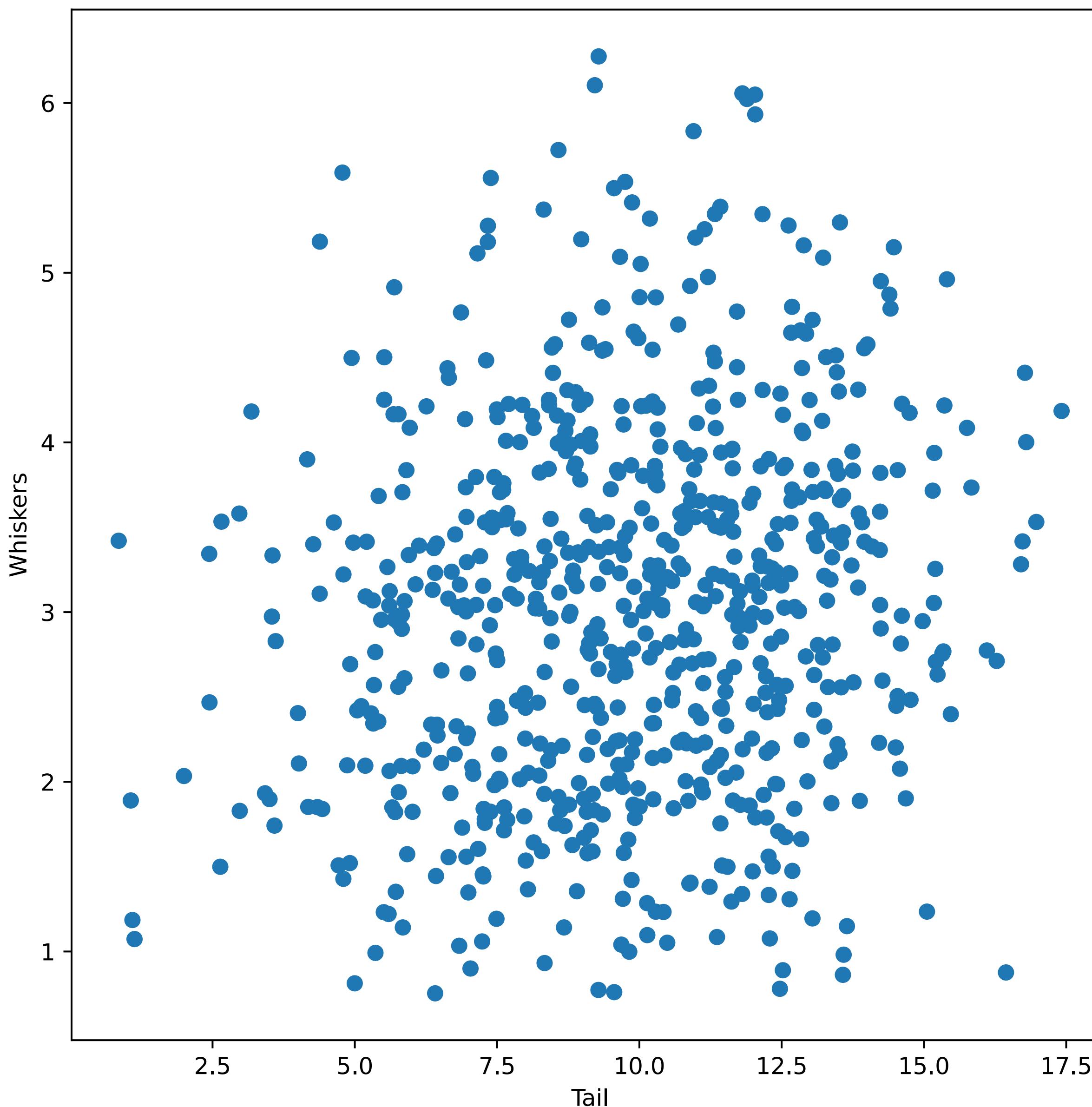


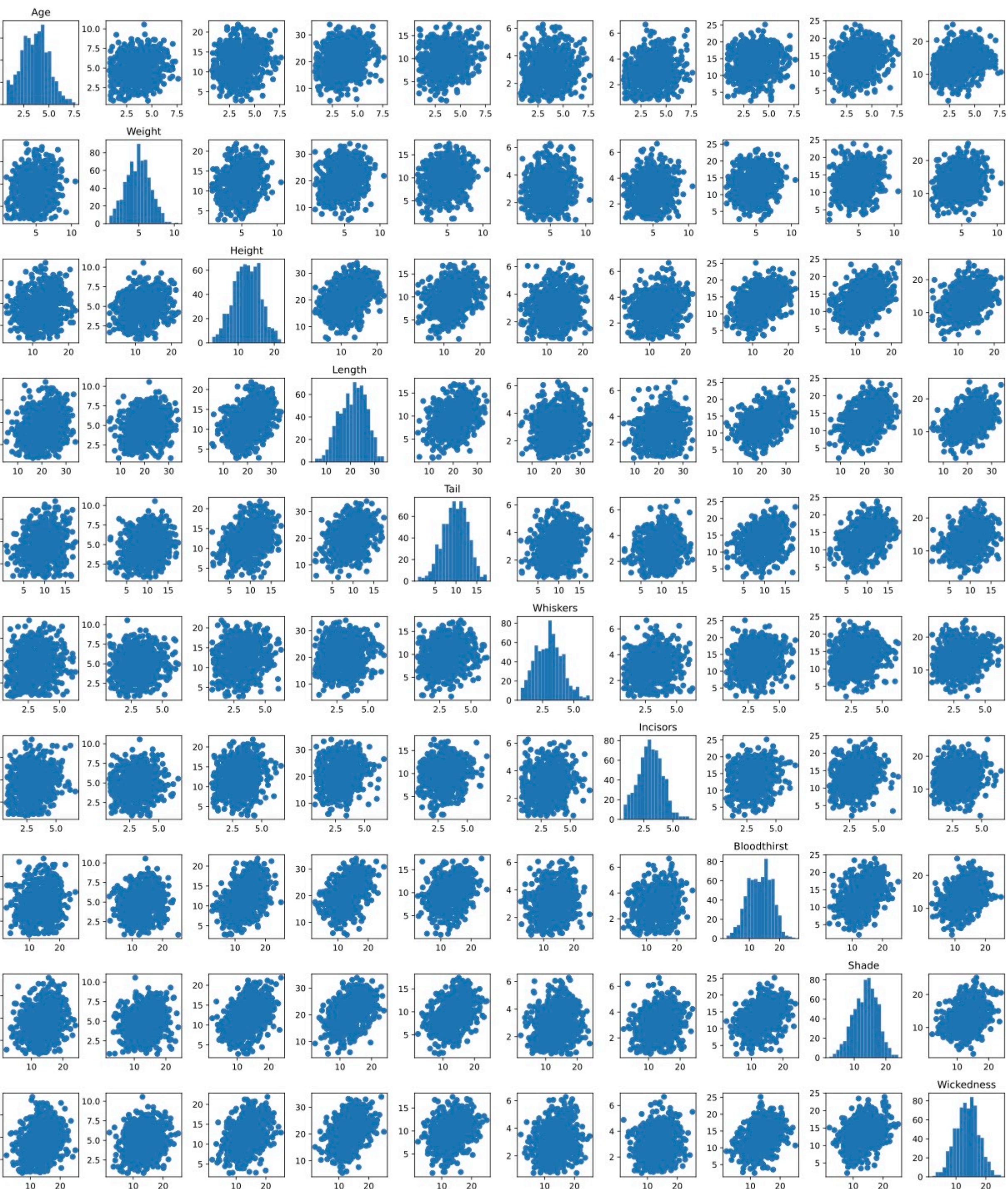
Age vs Weight





Tail vs Whiskers



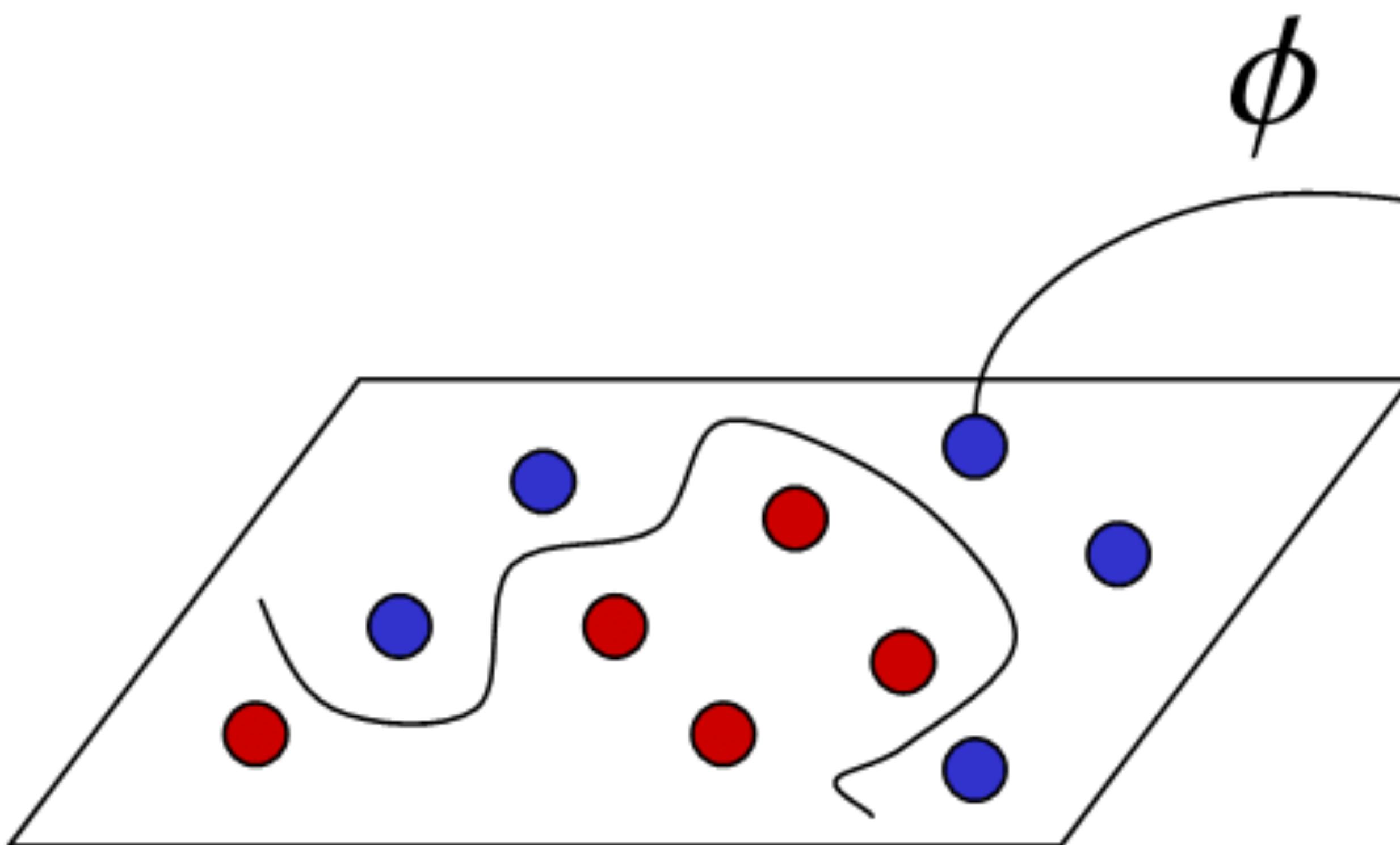




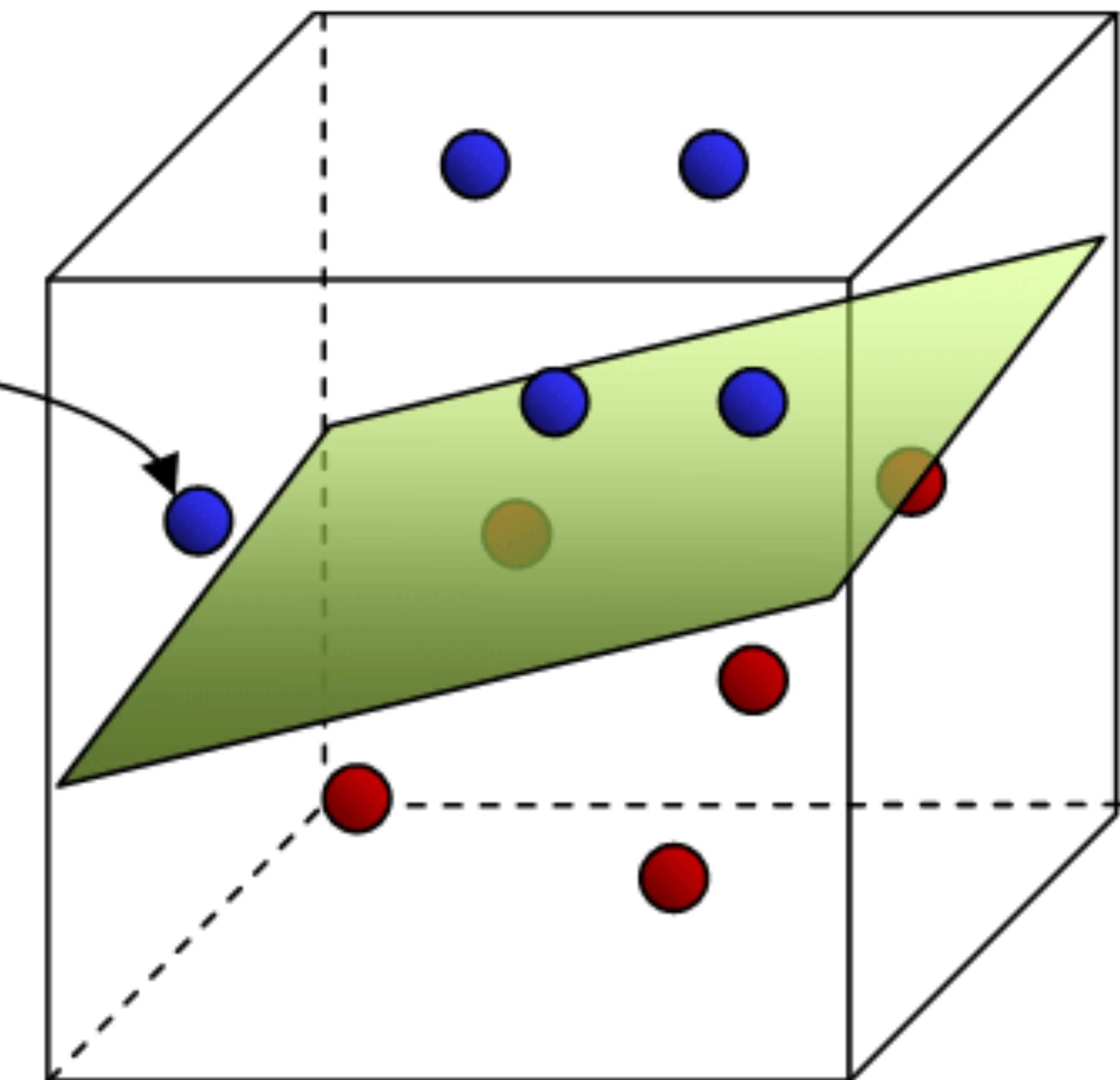




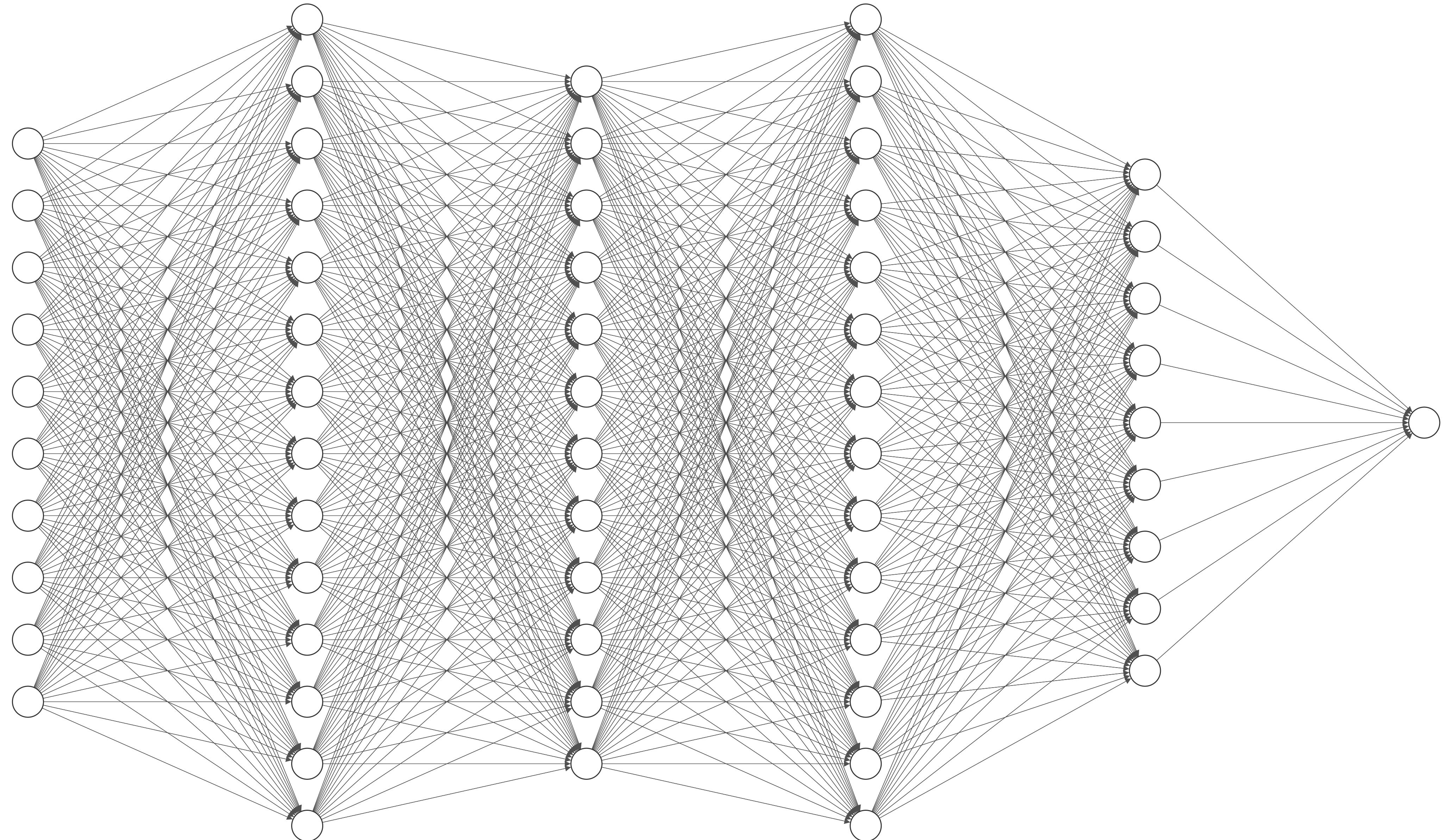
embeddings

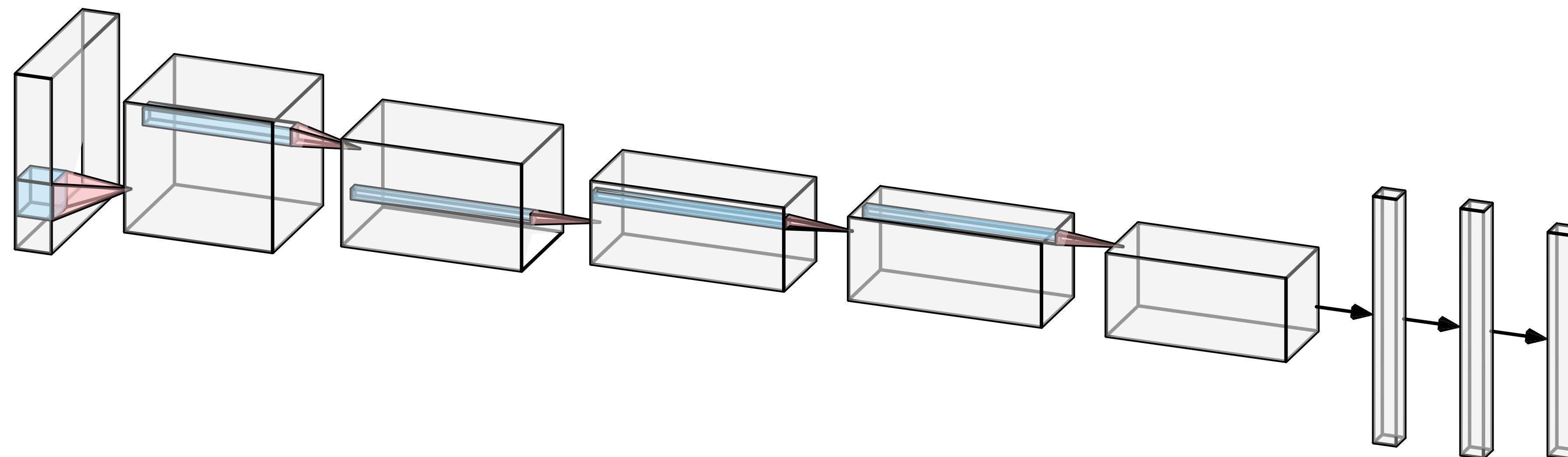


a) Input Space

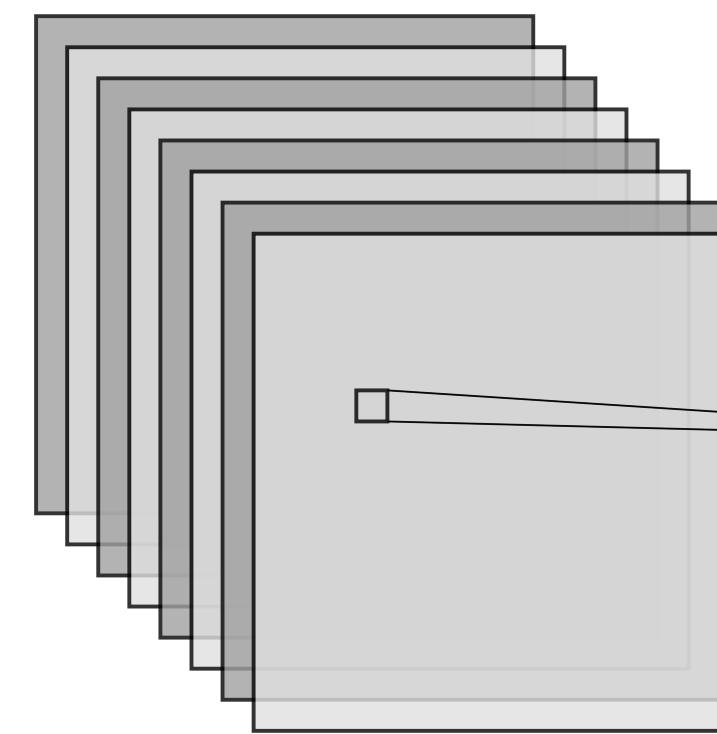


b) Feature Space

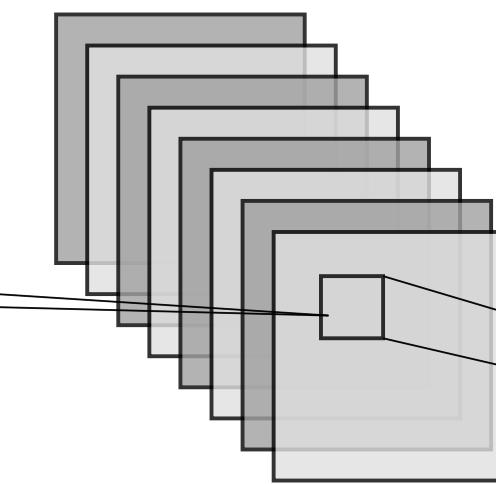




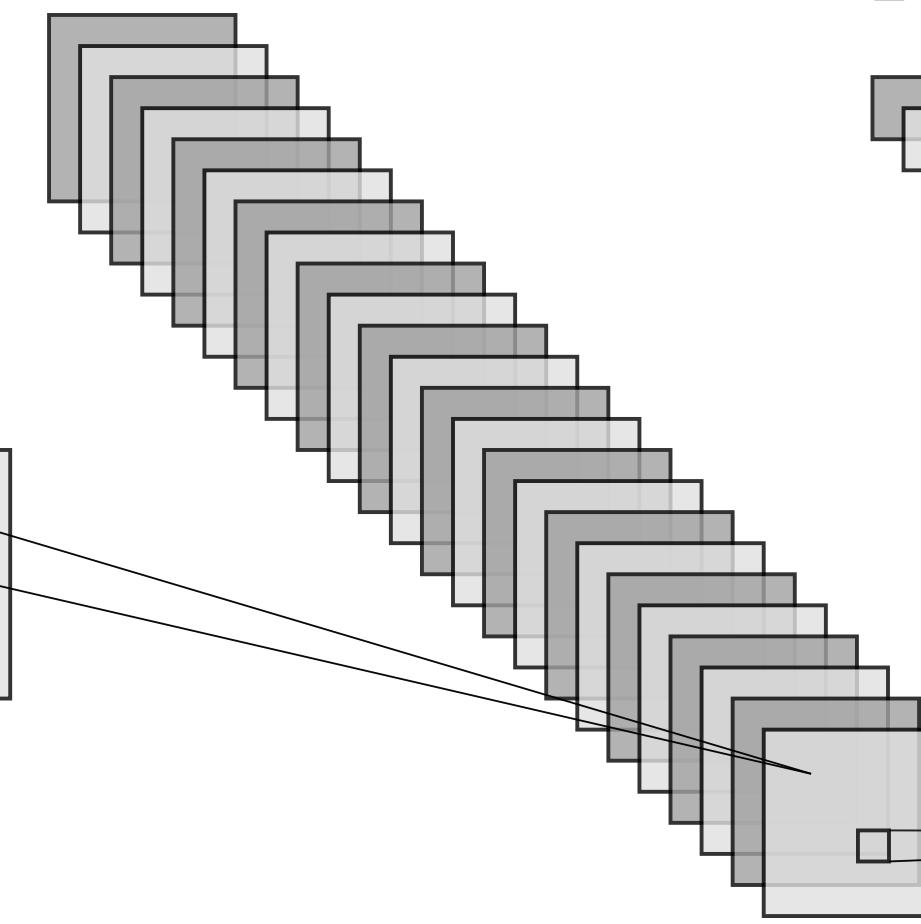
$8@128 \times 128$



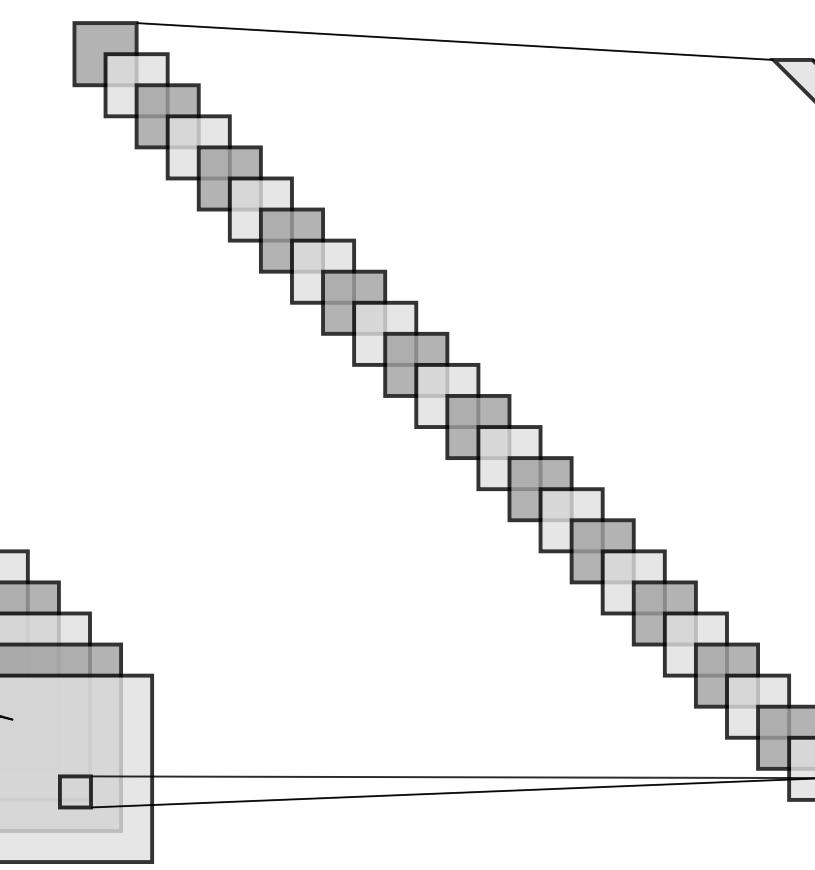
$8@64 \times 64$



$24@48 \times 48$



$24@16 \times 16$



1×256

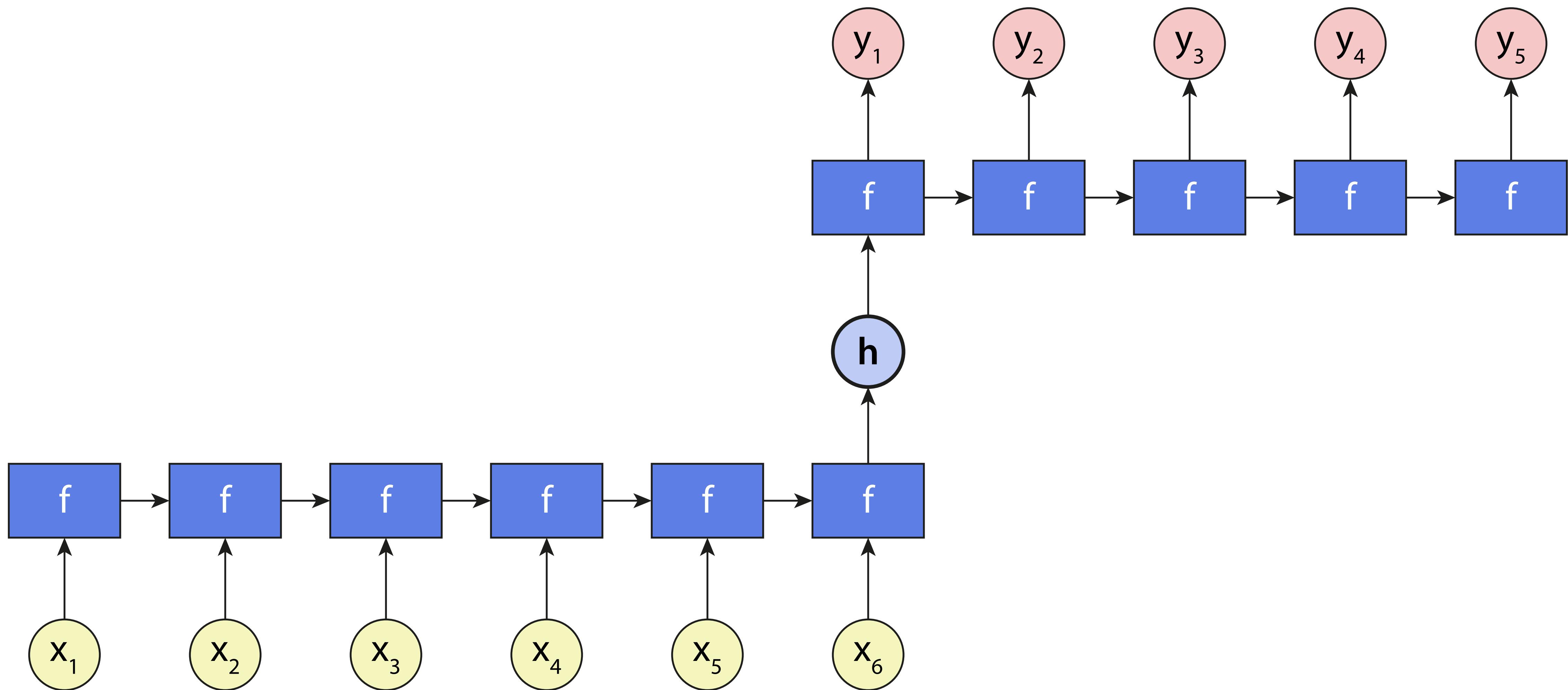
1×128

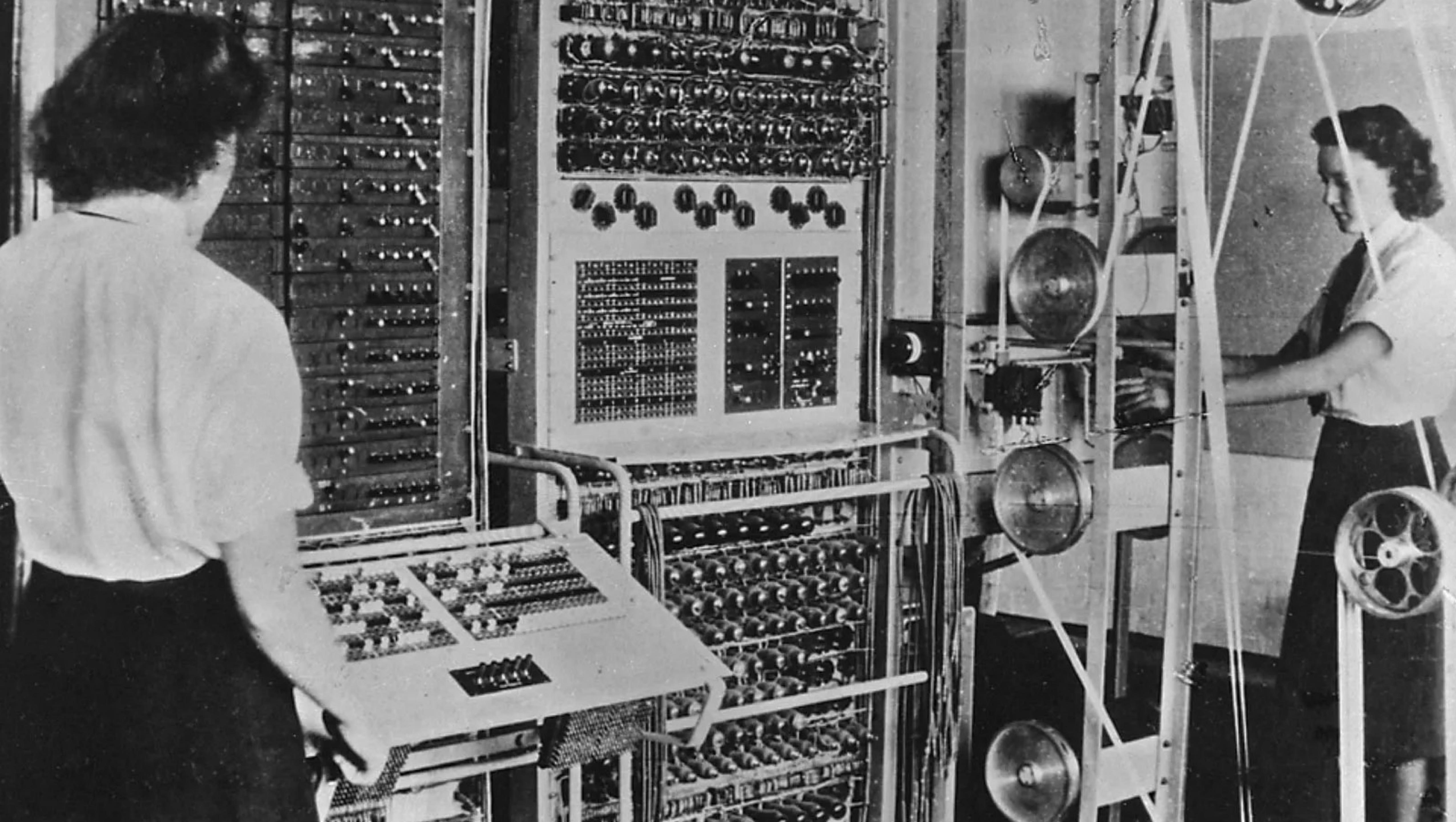
Max-Pool

Convolution

Max-Pool

Dense





00000110011001100110000110
0100000000000000100000000000
0000110011001100110000000000
0011001100110011001000000000
000000110011001100110011001100
11001000000000000000000011001100
100111001000000000000000000000001

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

x^T

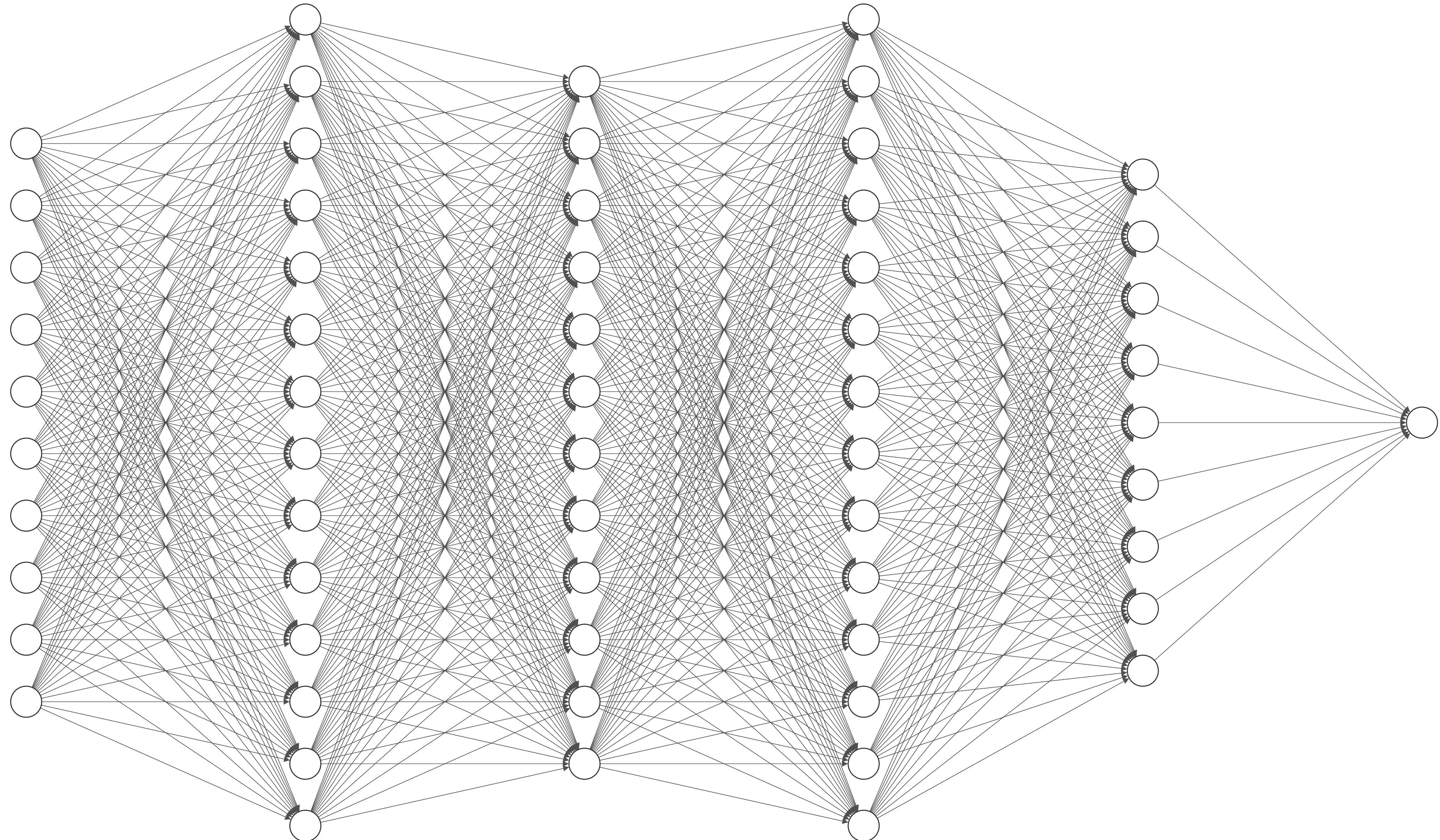
(one-hot)

 $[0, 1, 0, 0, 0] \times$ W \tilde{x}^T

(embedding)

	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.)

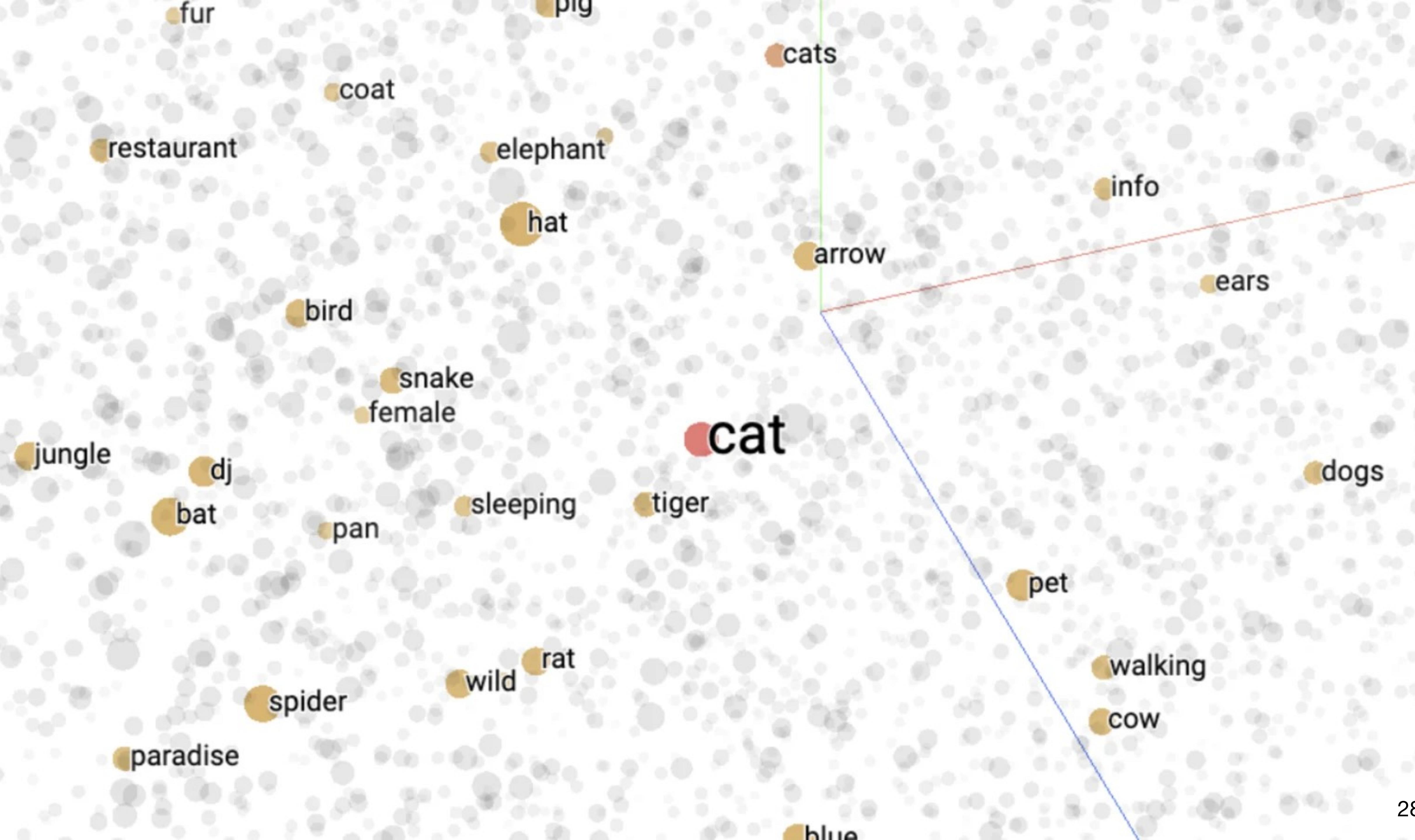
I did not sleep well, though my bed was comfortable enough, for I had all sorts of queer dreams. There was a dog howling all night under my window, which may have had something to do with it; or it may have been the paprika, for I had to drink up all the water in my carafe, and was still thirsty. Towards 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?

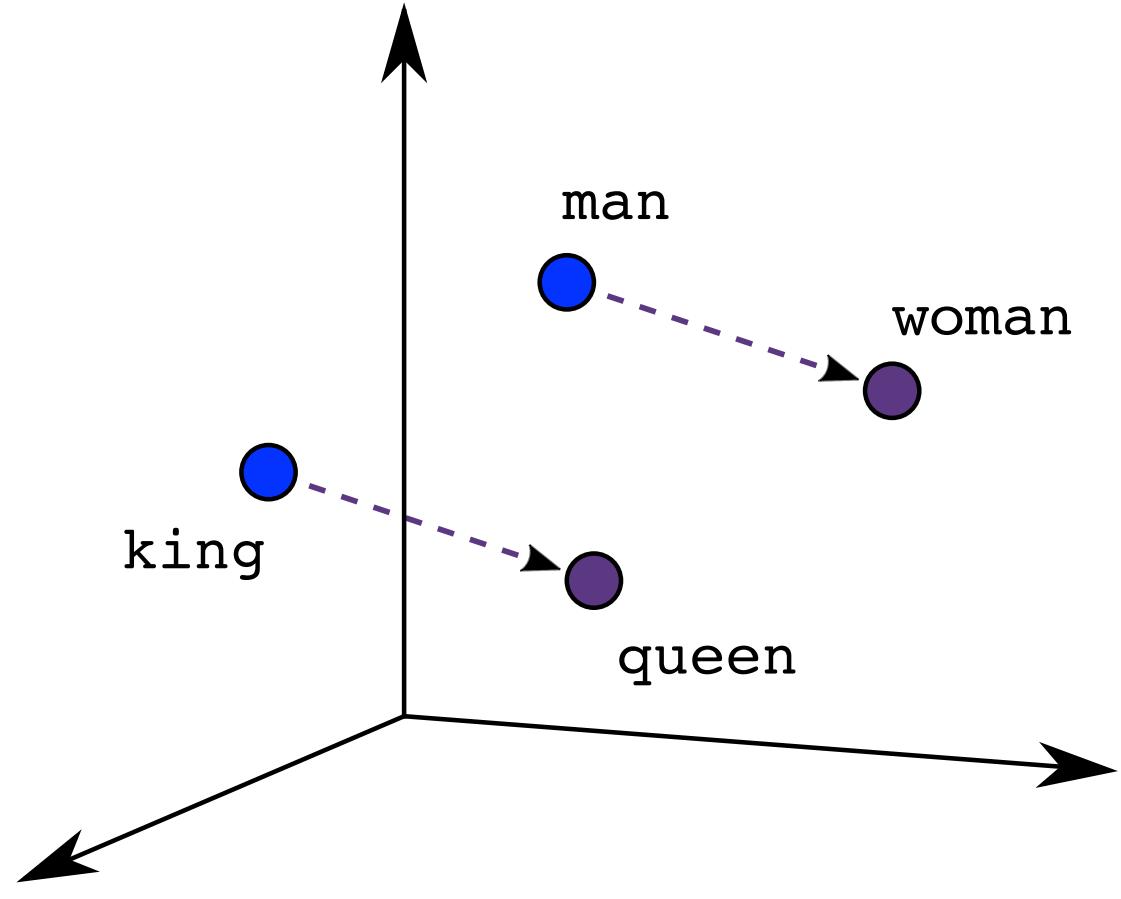
All day long we seemed to dawdle through a country which was full of beauty of every kind. Sometimes we saw little towns or castles on the top of steep hills such as we see in old missals; sometimes we ran by rivers and streams which seemed from the wide stony margin on each side of them to be subject 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.

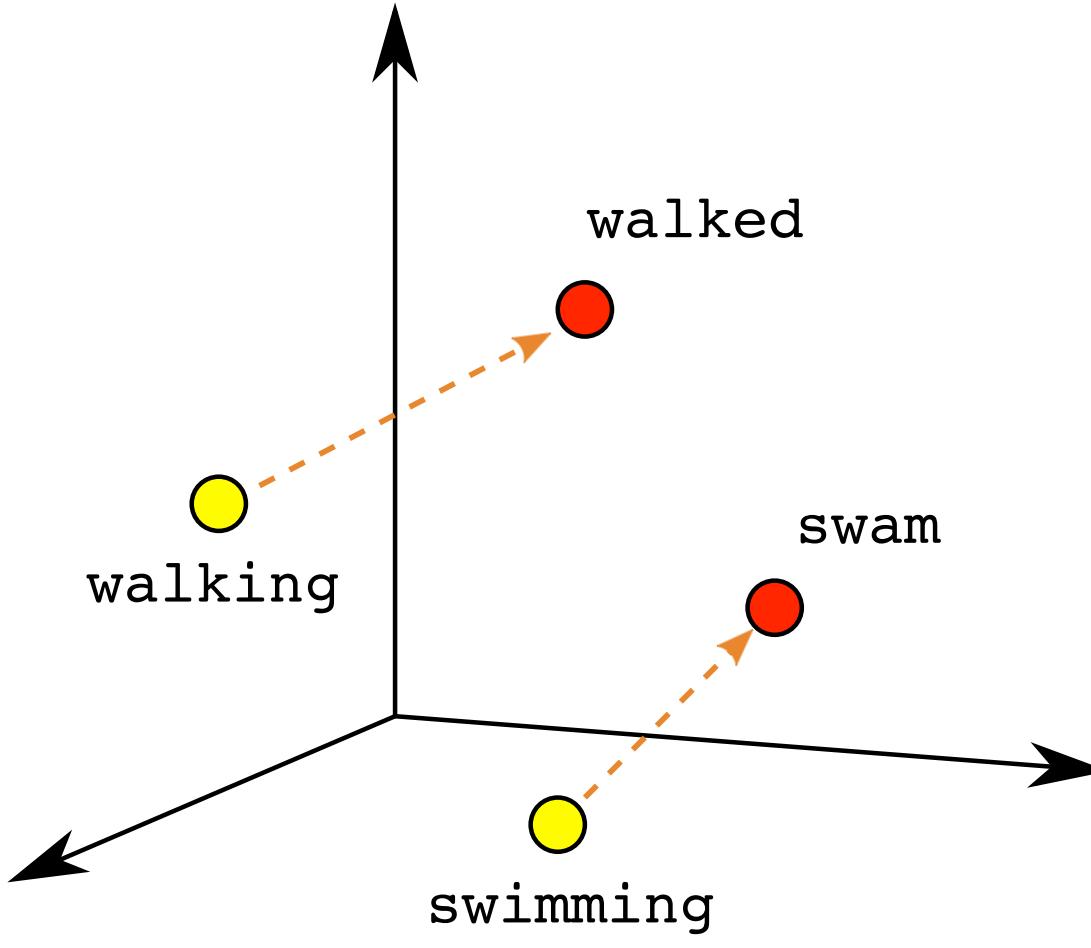
Count Dracula had directed me to go to the Golden Krone Hotel, which I found, to my great delight, to be thoroughly old-fashioned, for of course I wanted to see all I could of the ways of the country. I was evidently expected, for when I got near the door I faced a cheery-looking elderly woman in the 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."

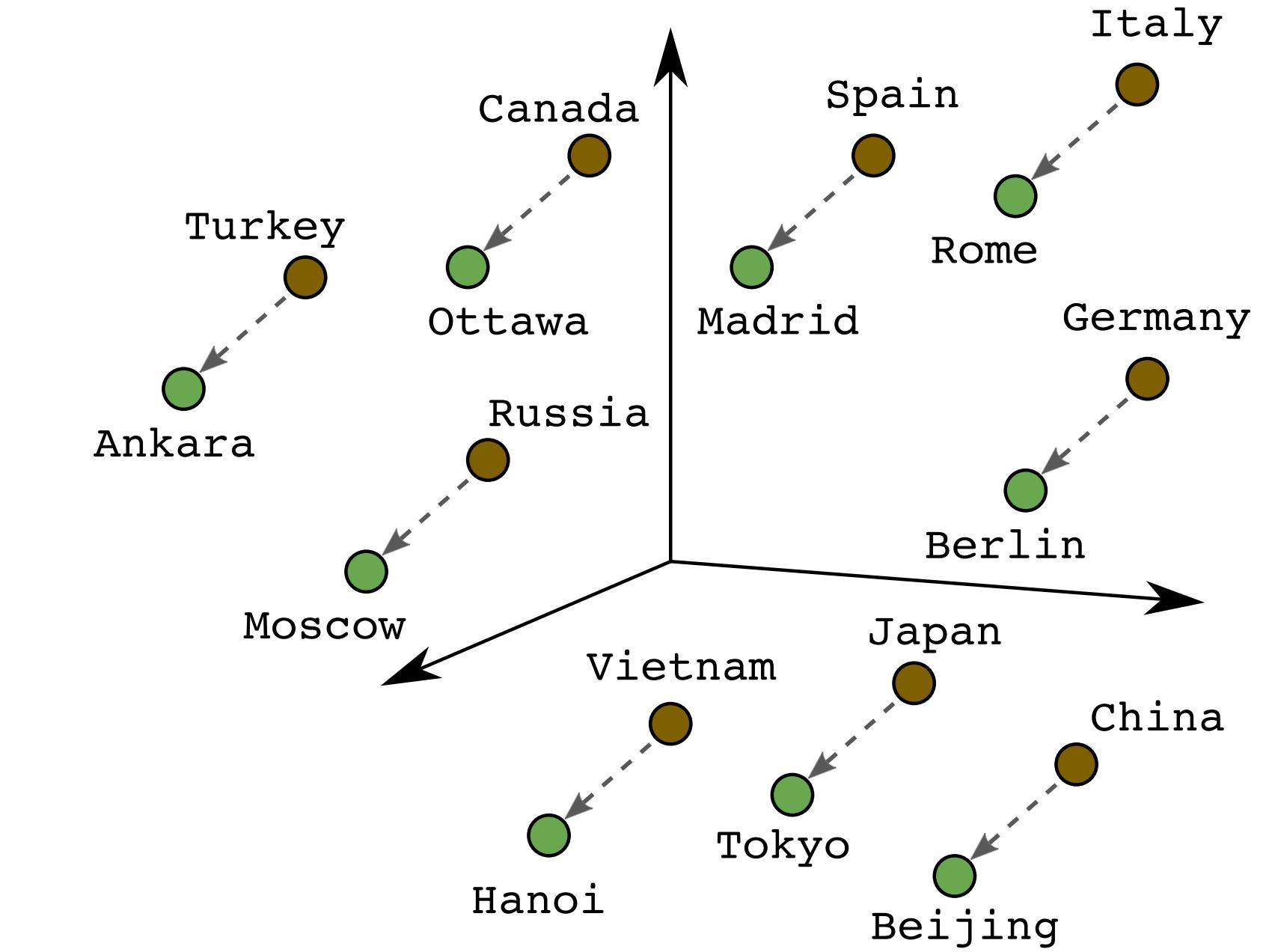




Male-Female



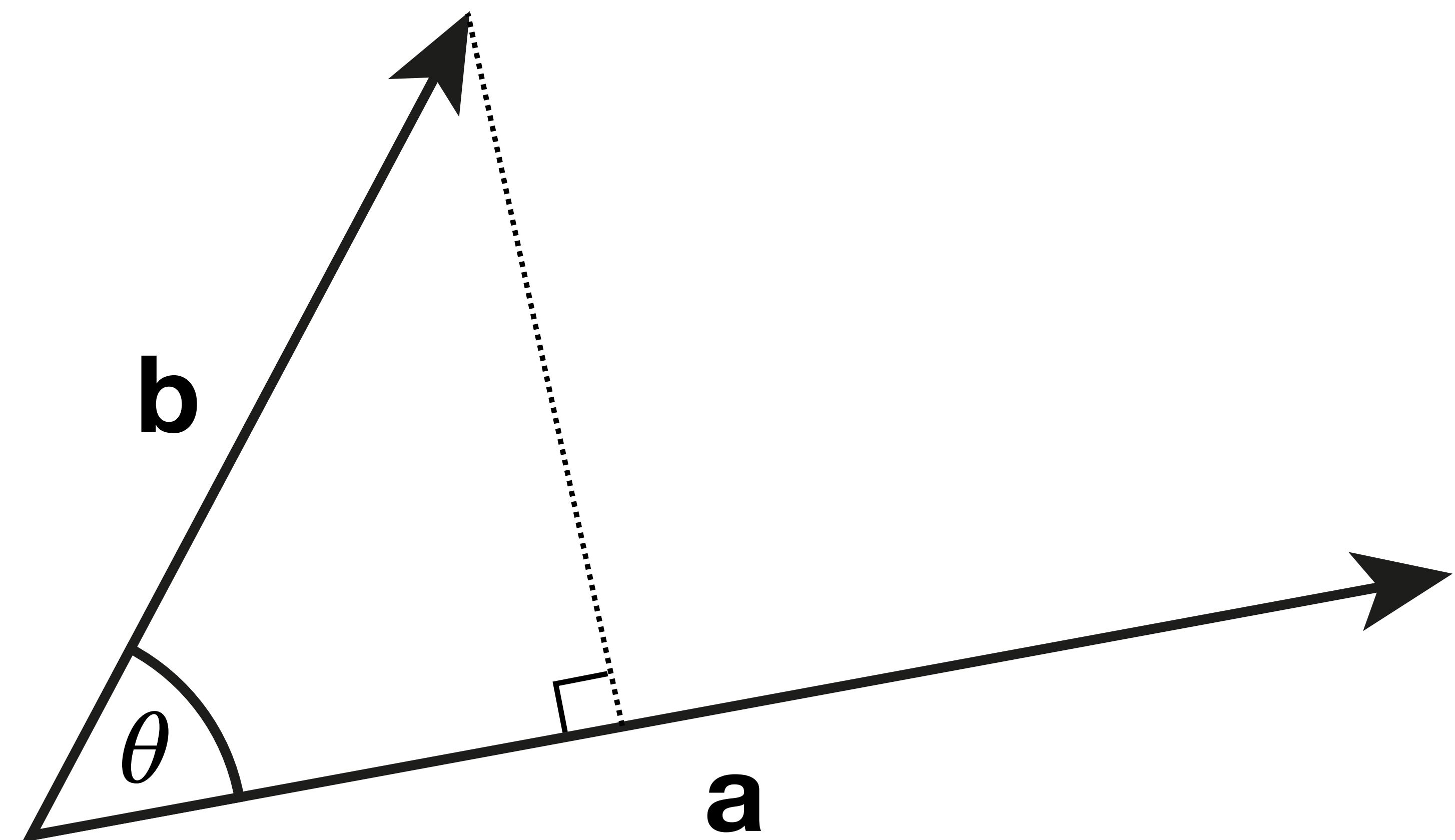
Verb Tense

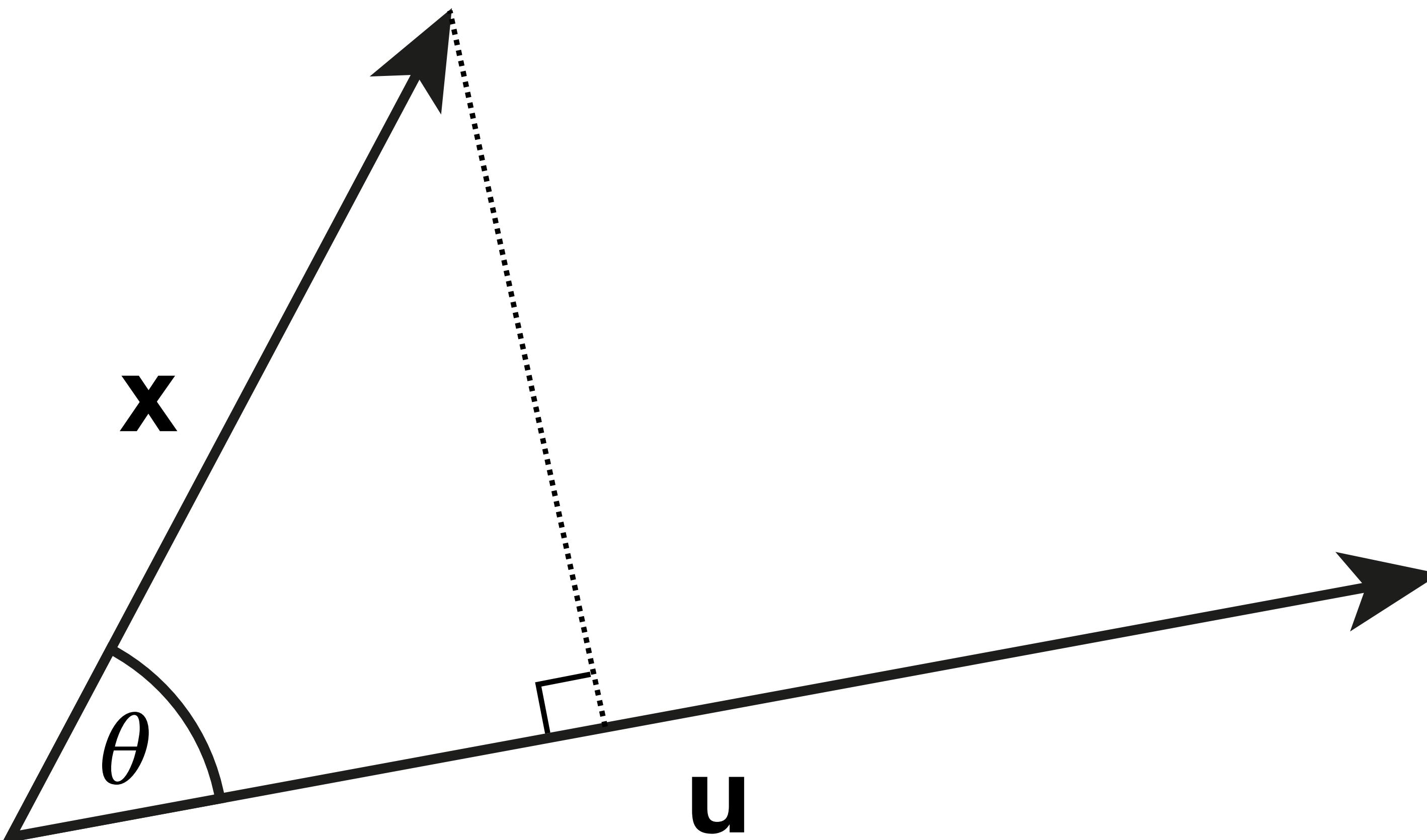


Country-Capital

PCA

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





$$|u| = 1$$

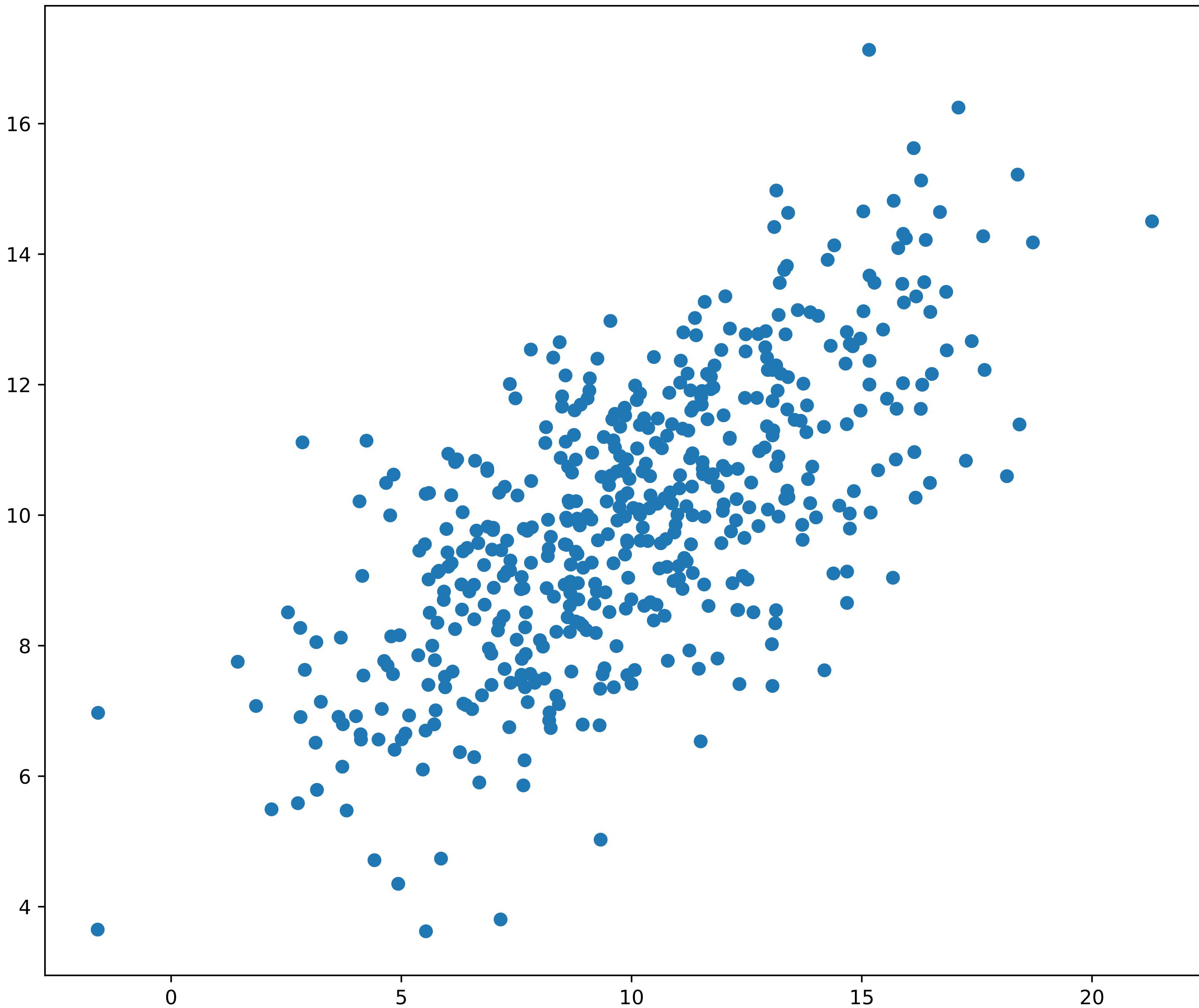
$$\mathbf{x} \cdot \mathbf{u} = |\mathbf{x}| \cos \theta$$

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

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

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

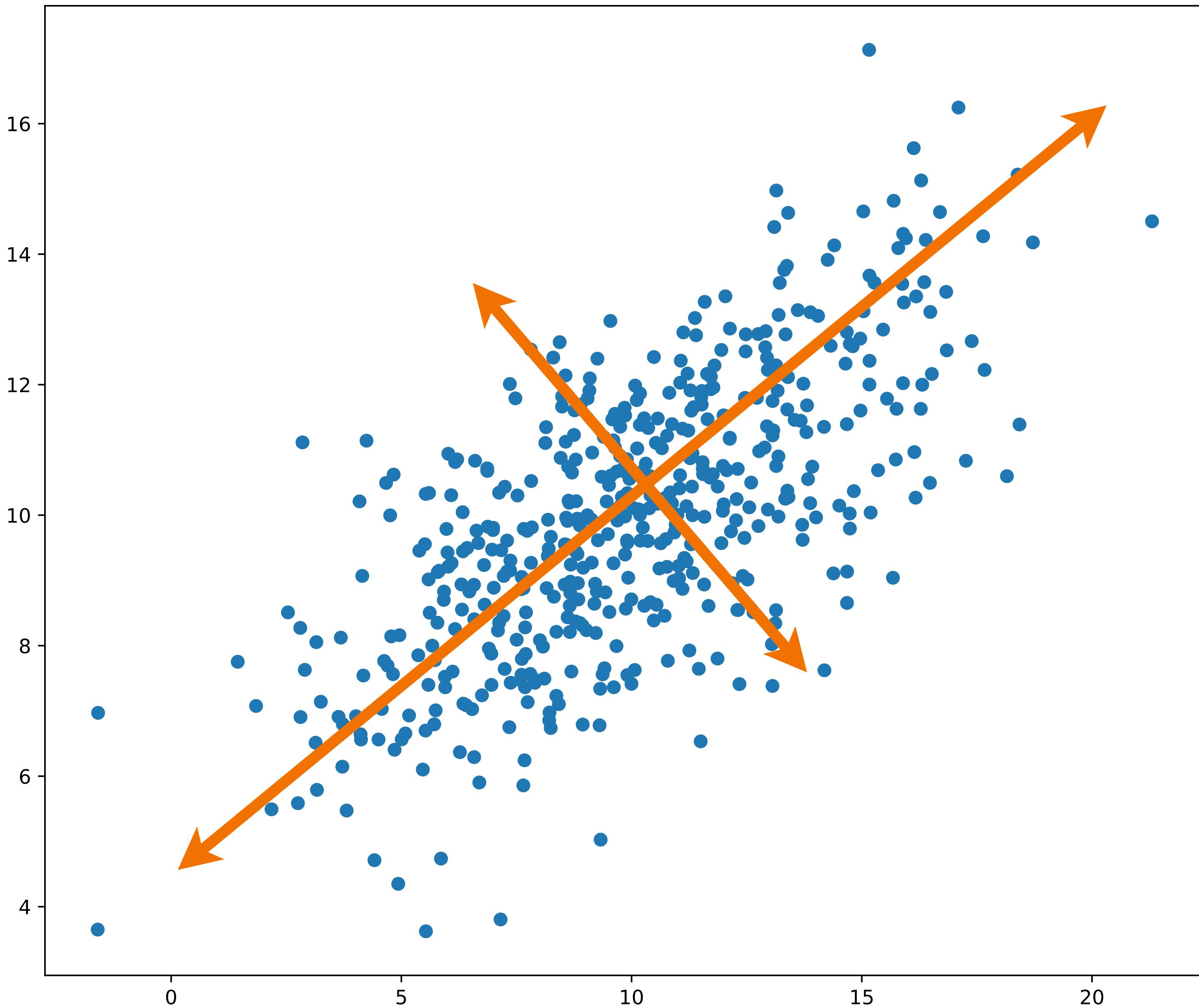
$\tilde{\mathbf{x}} = \mathbf{xu}$

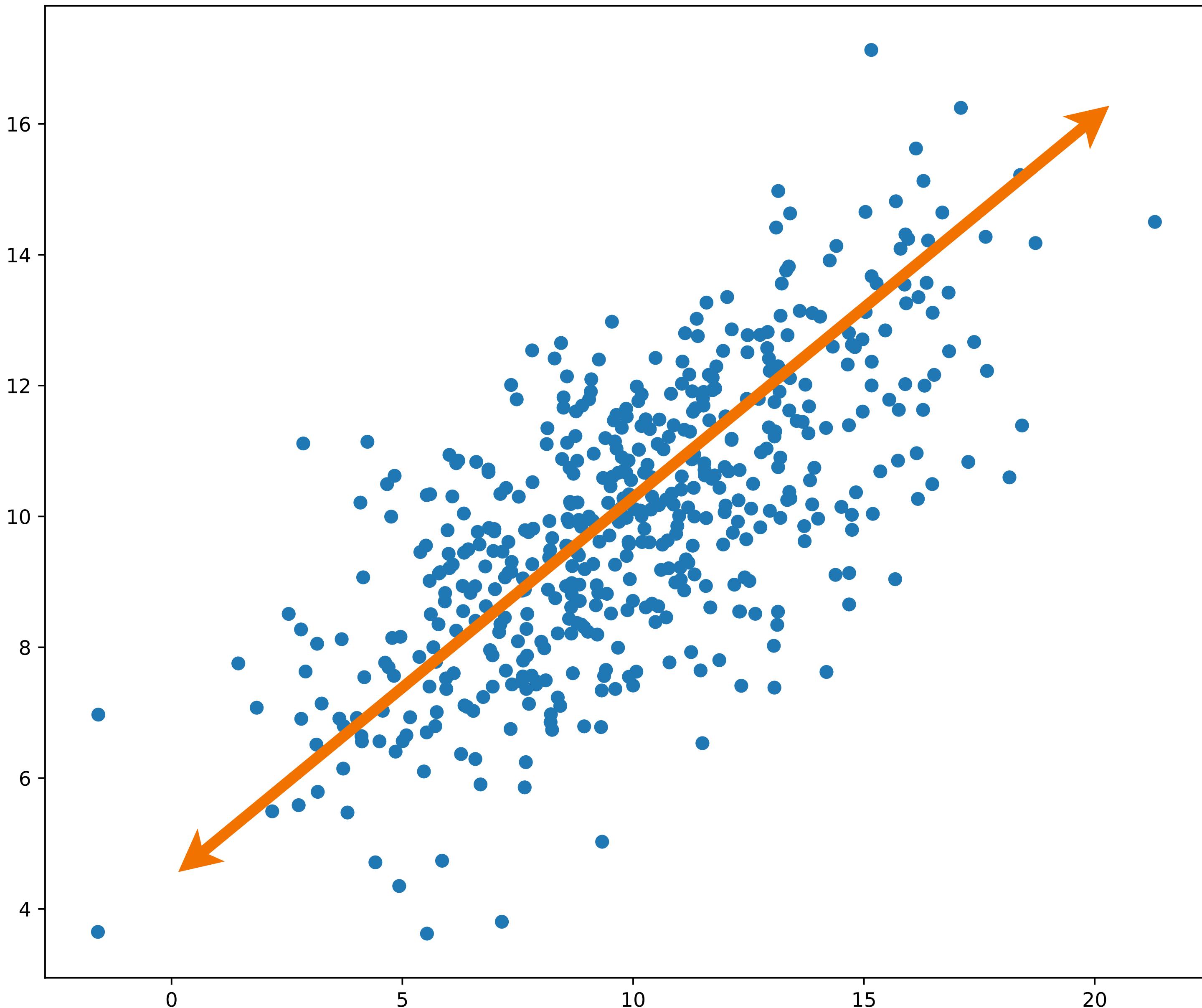


$$\Sigma = \begin{bmatrix} \text{var}(\mathbf{x}_{|1}) & \text{covar}(\mathbf{x}_{|1}, \mathbf{x}_{|2}) & \cdots & \text{covar}(\mathbf{x}_{|1}, \mathbf{x}_{|d}) \\ \text{covar}(\mathbf{x}_{|2}, \mathbf{x}_{|1}) & \text{var}(\mathbf{x}_{|2}) & \cdots & \text{covar}(\mathbf{x}_{|2}, \mathbf{x}_{|d}) \\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

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

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





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

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

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

maximise $\mathbf{u}^T \Sigma \mathbf{u}$
 \mathbf{u}

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

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

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

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

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

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

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

$$\sqrt{\text{diag}(\text{eig}(\mathbf{X}\mathbf{X}^\top))}$$

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

cols eigenvectors of $\mathbf{X}^\top\mathbf{X}$

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

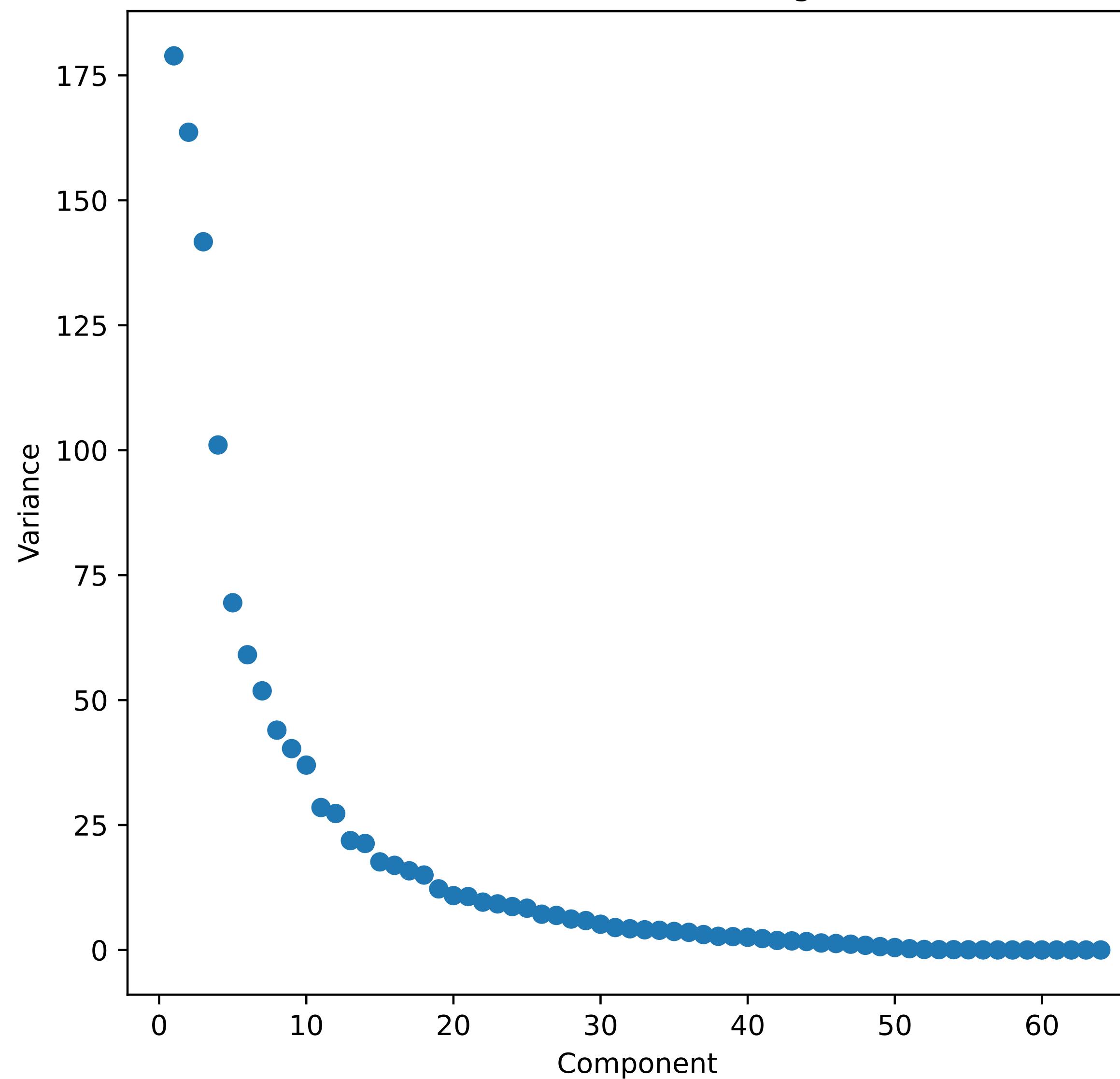
$$\tilde{\mathbf{X}} = \mathbf{X}\mathbf{U}$$

$$= \mathbf{V}\mathbf{D}\mathbf{U}^T\mathbf{U}$$

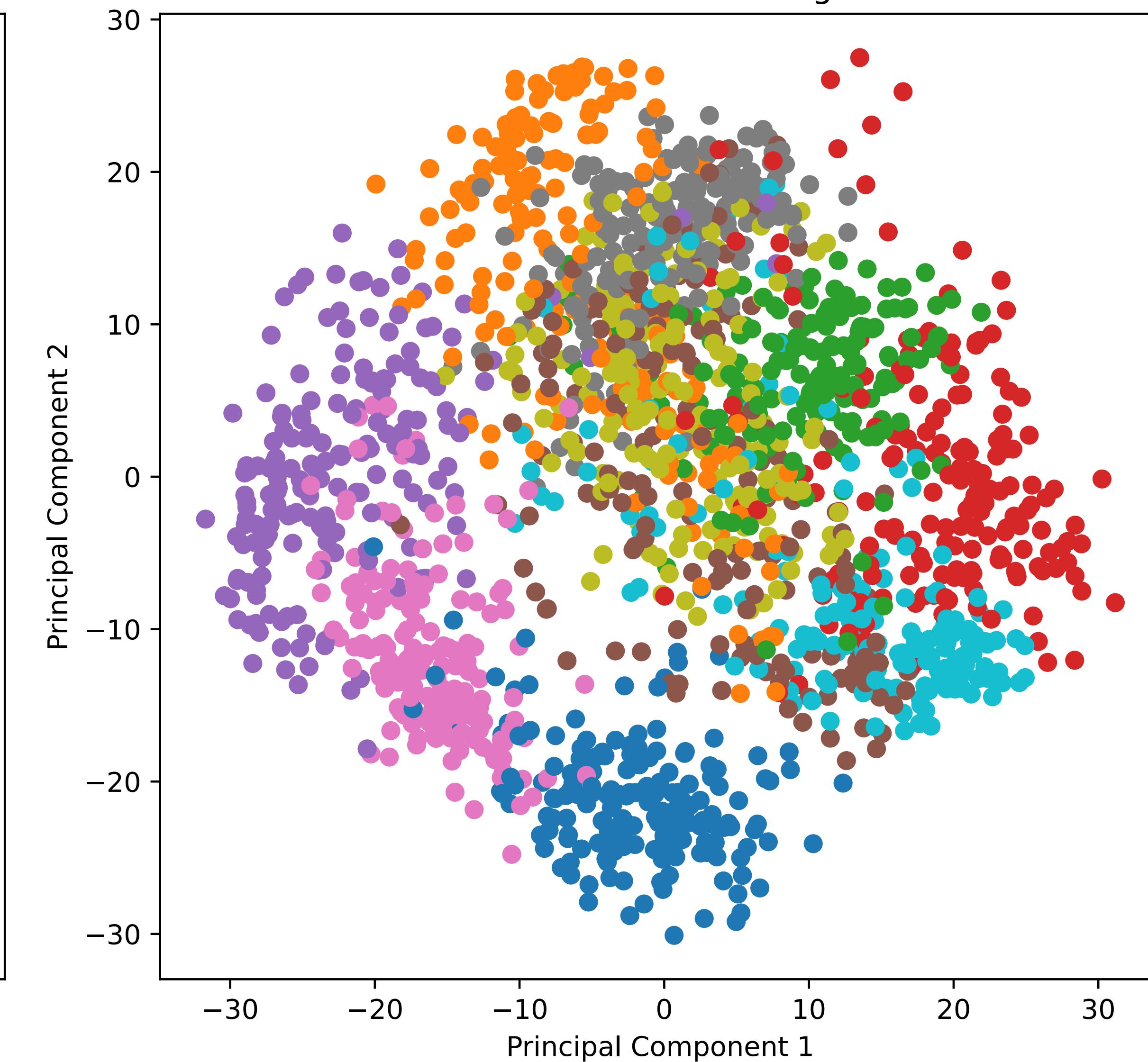
$$= \mathbf{V}\mathbf{D}$$

8 907390125
6 10673760748
7 1009737630139
1 9976074750925
6 10009737630139
1 9976074750925
3 10009737630139
6 10009737630139
1 9976074750925
6 10009737630139
1 9976074750925

PCA of Handwritten Digits

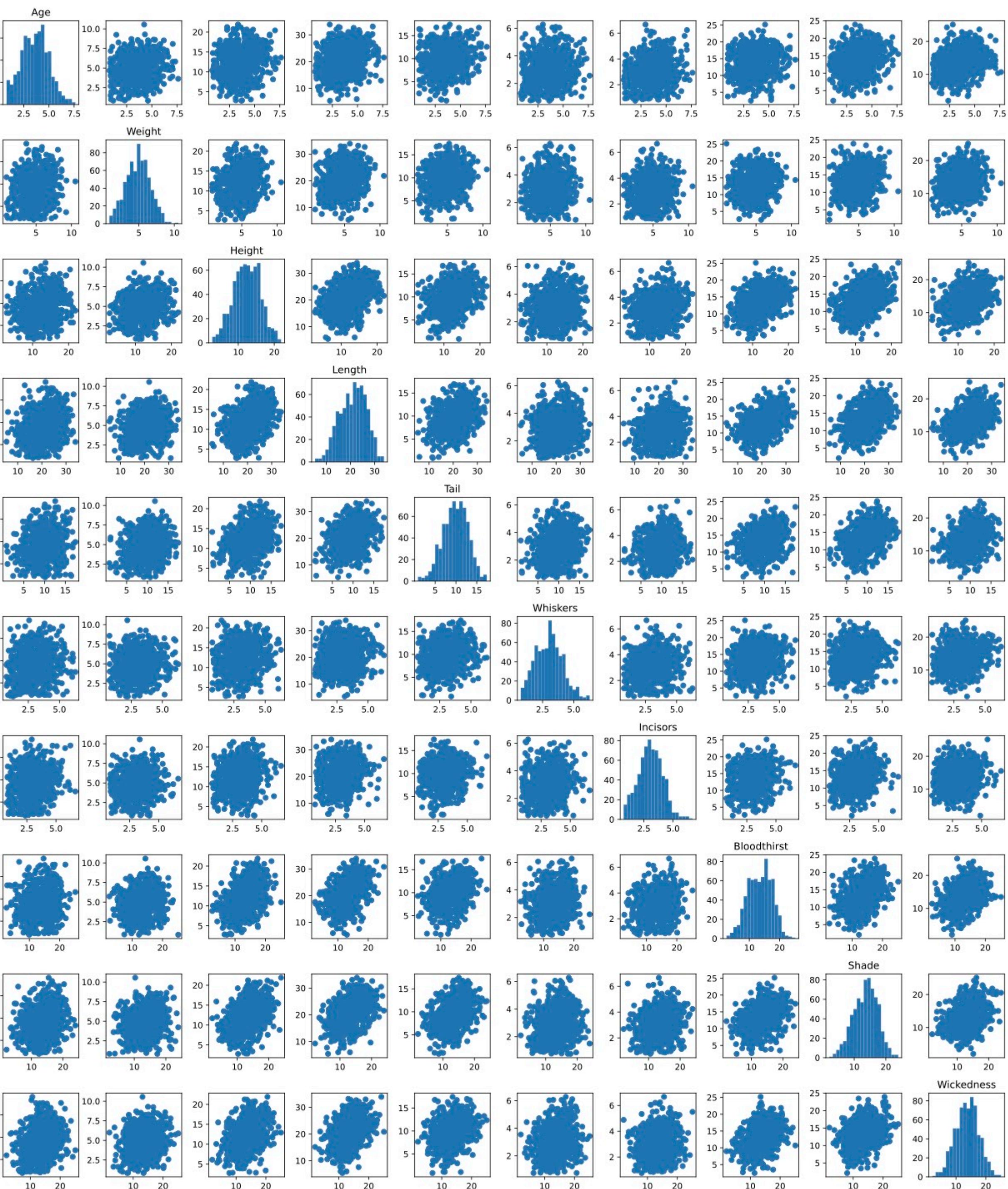


PCA of Handwritten Digits

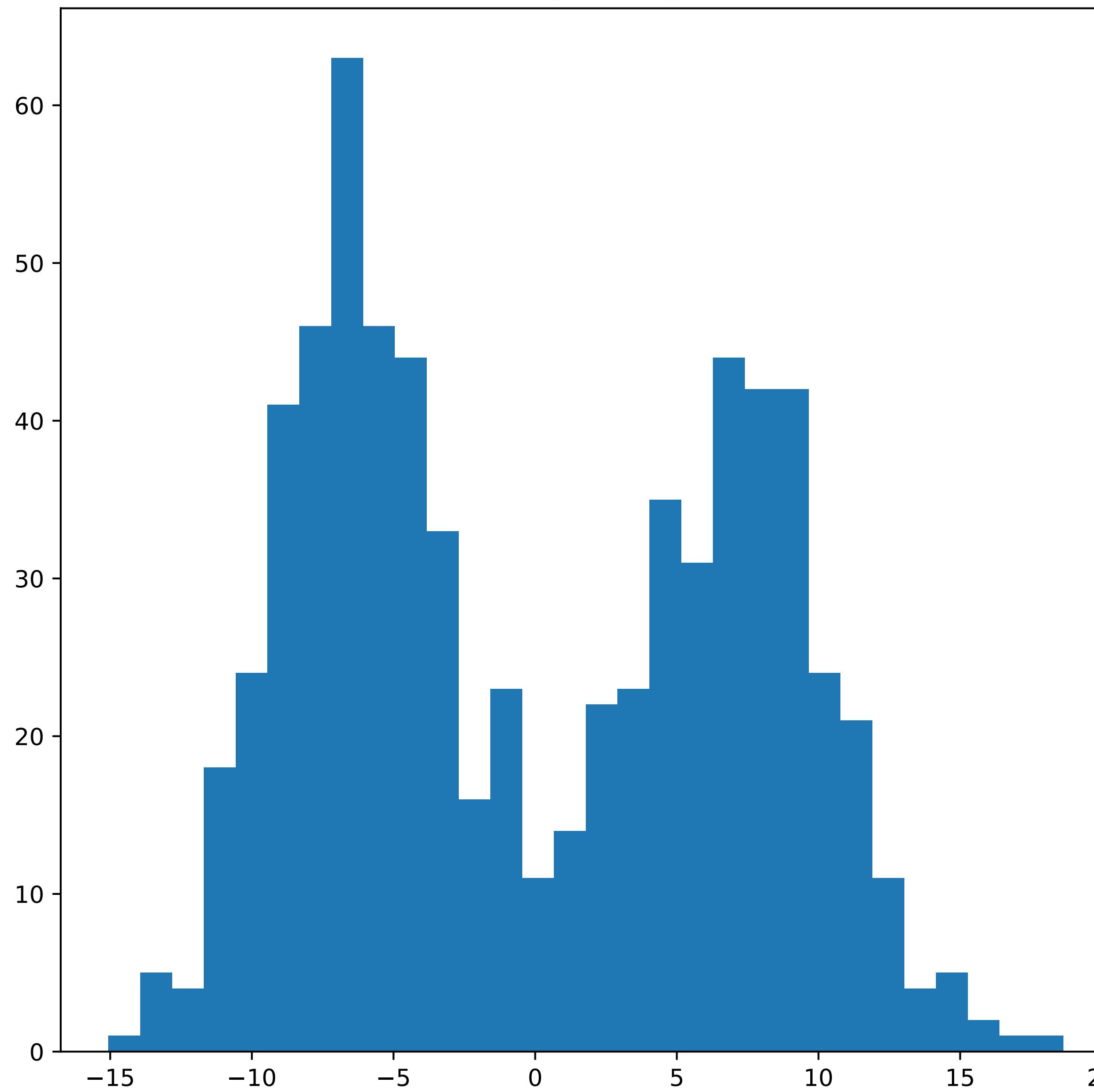




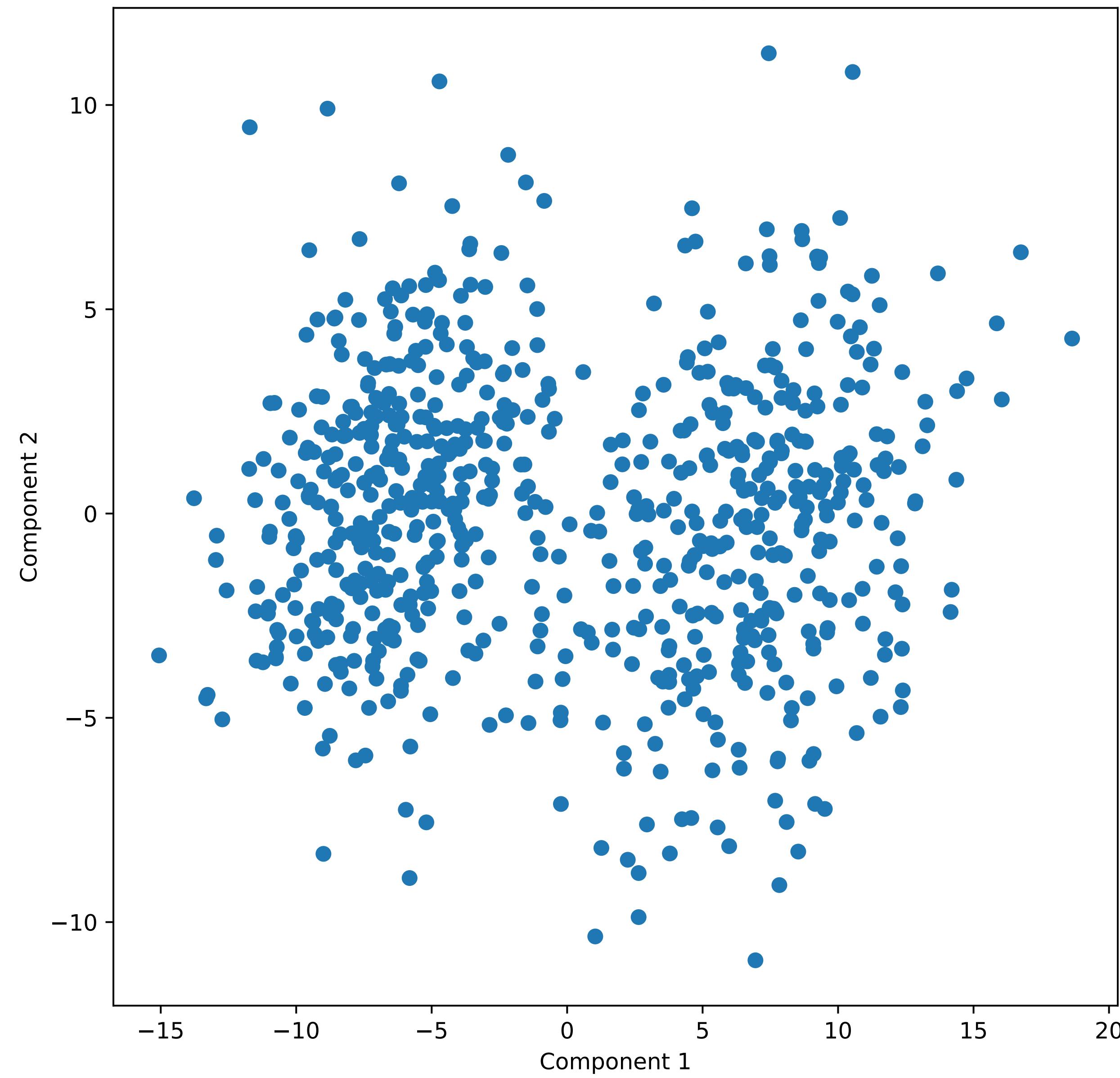




First Principal Component



First 2 Principal Components



clustering

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

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

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

Initialise c somehow

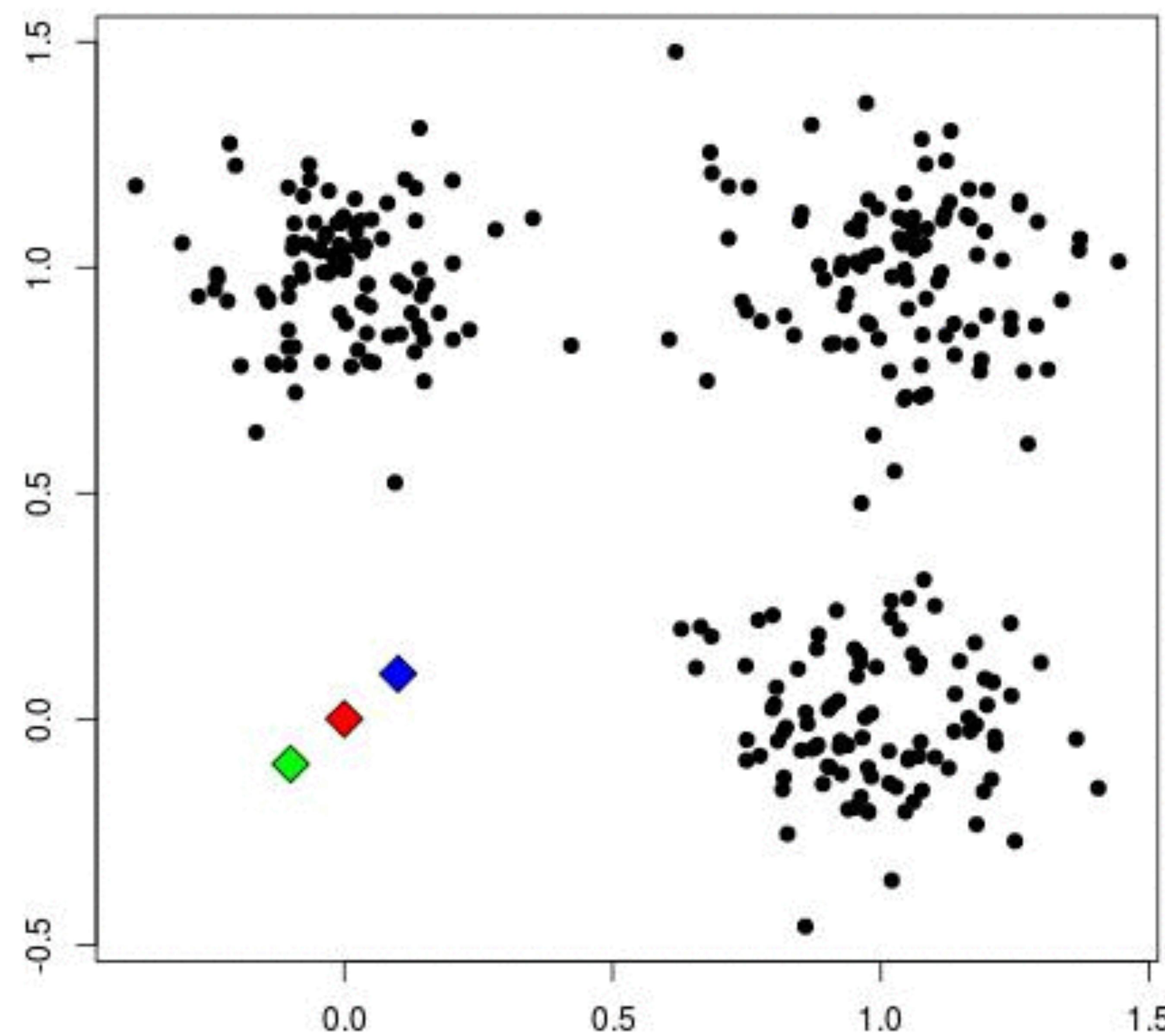
Repeat

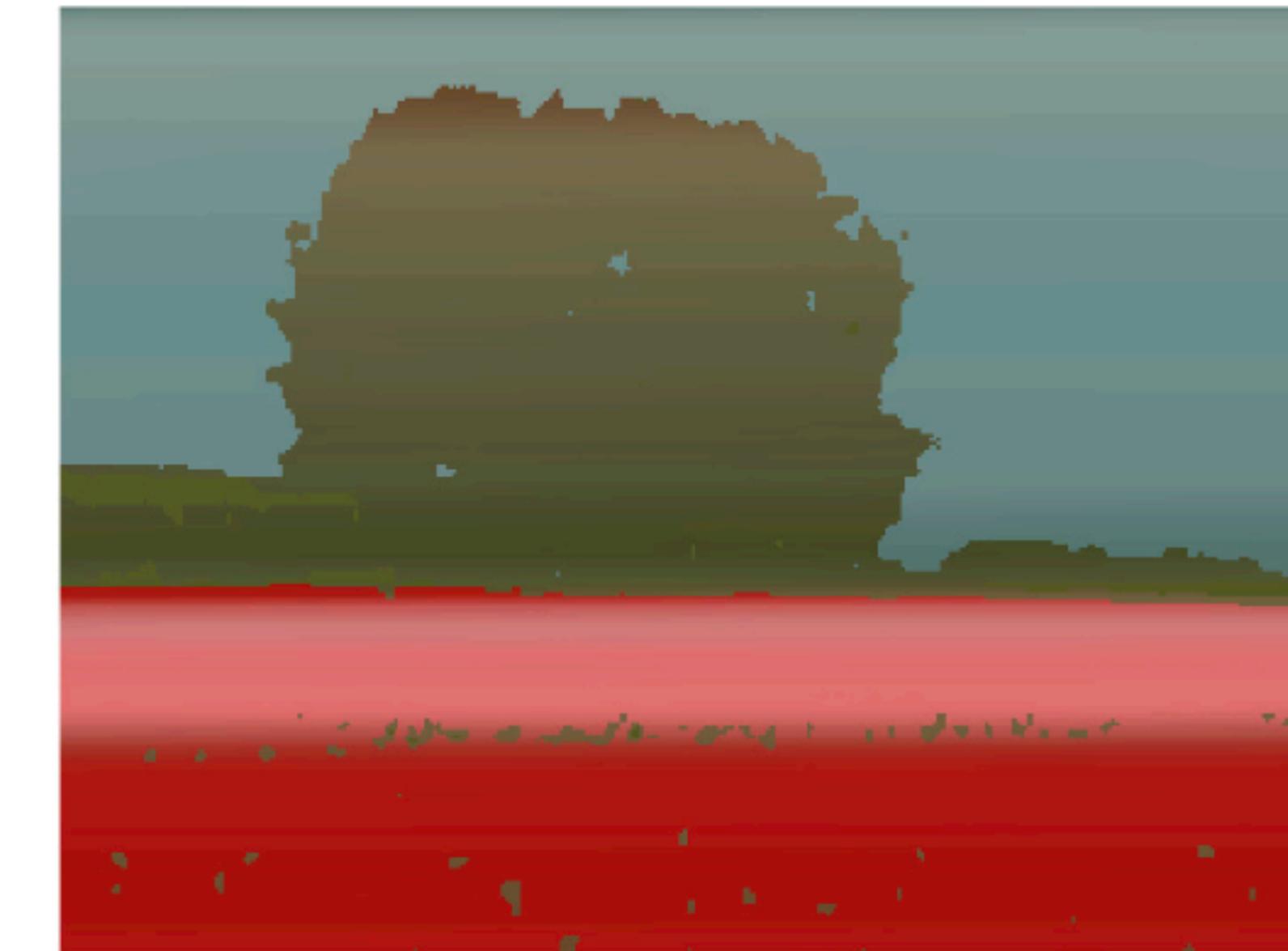
 Assign each x to its nearest c

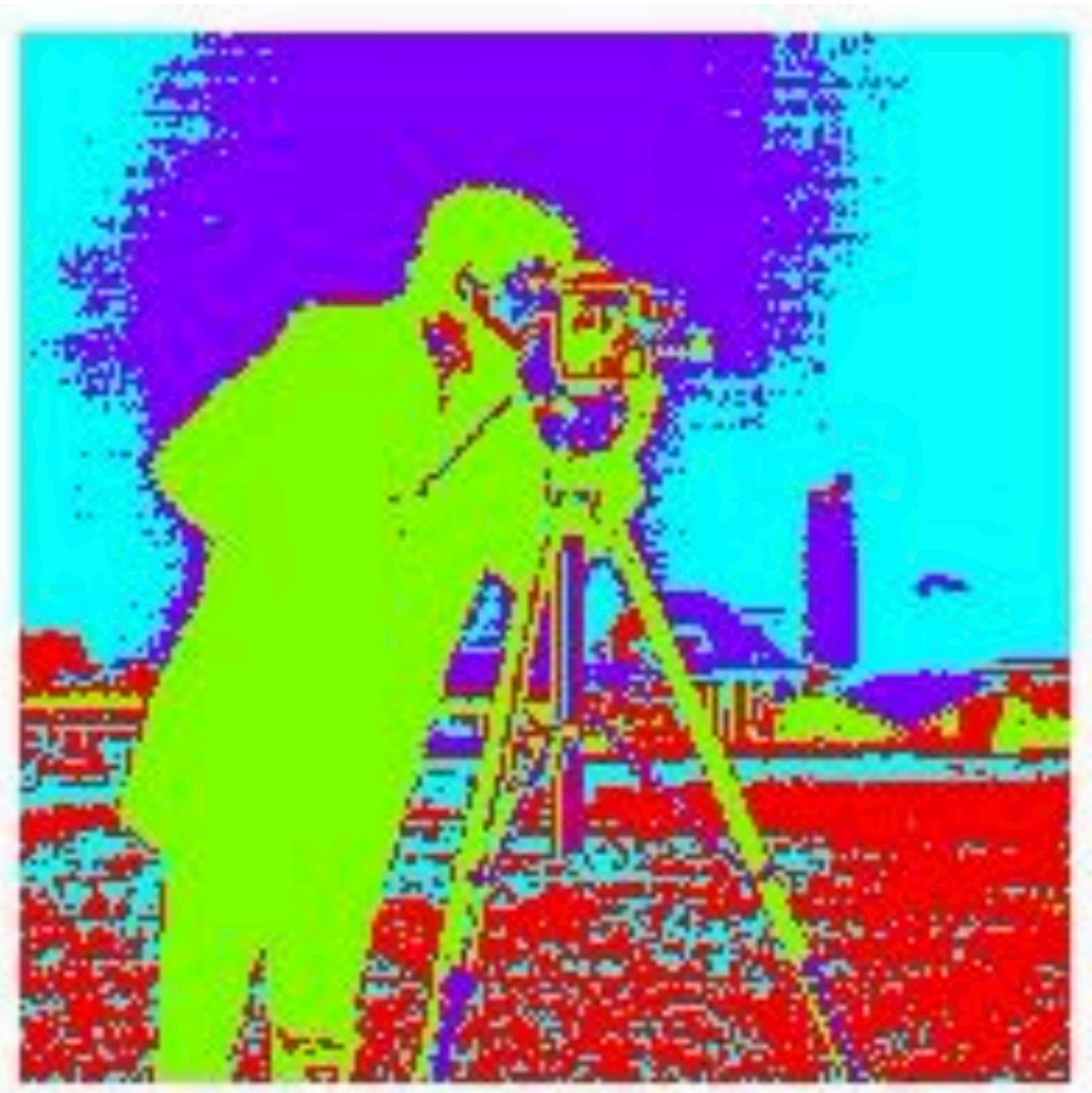
 Update each c to be the mean of all x assigned to it

Until nothing changes

Start!

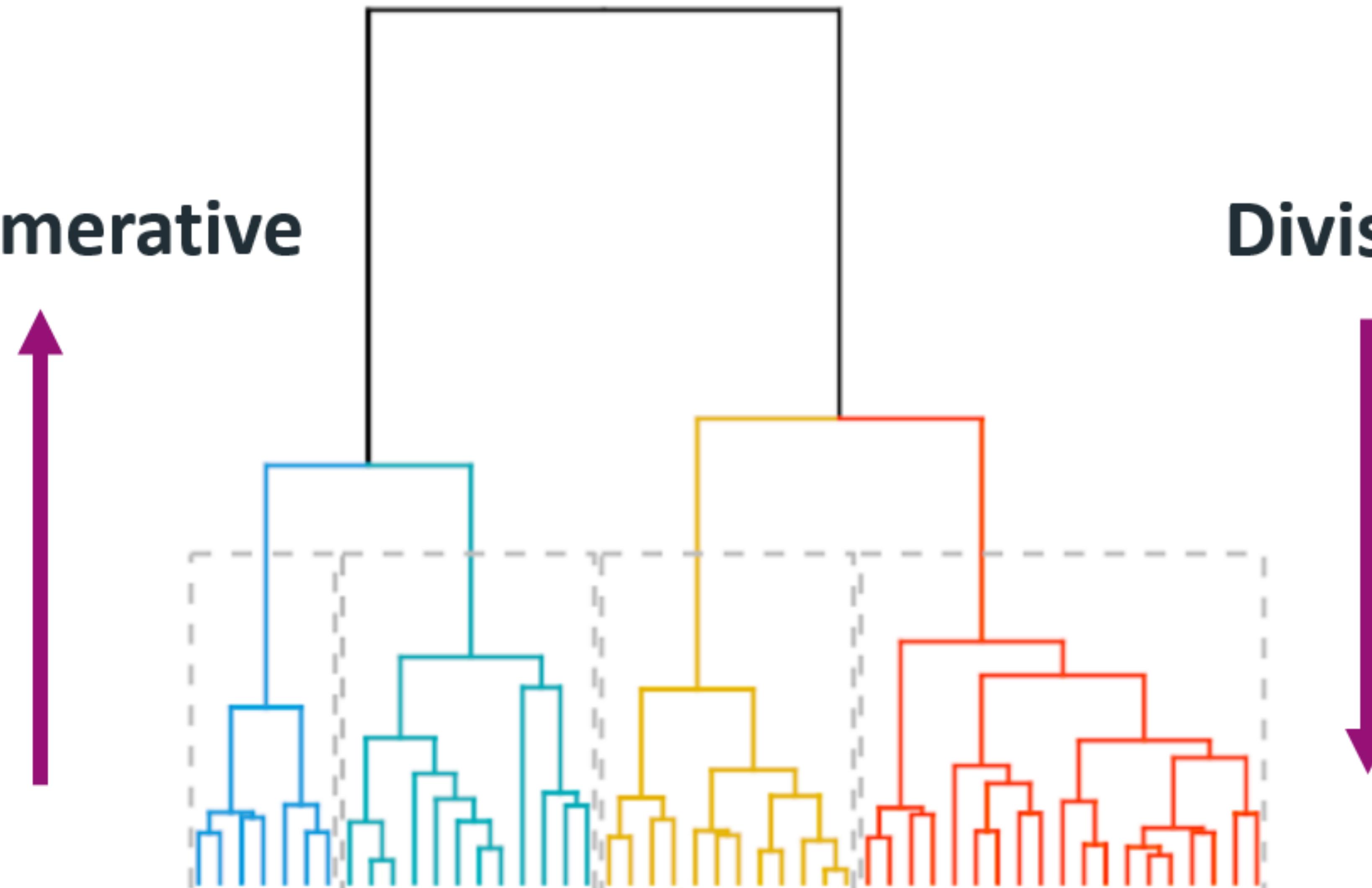






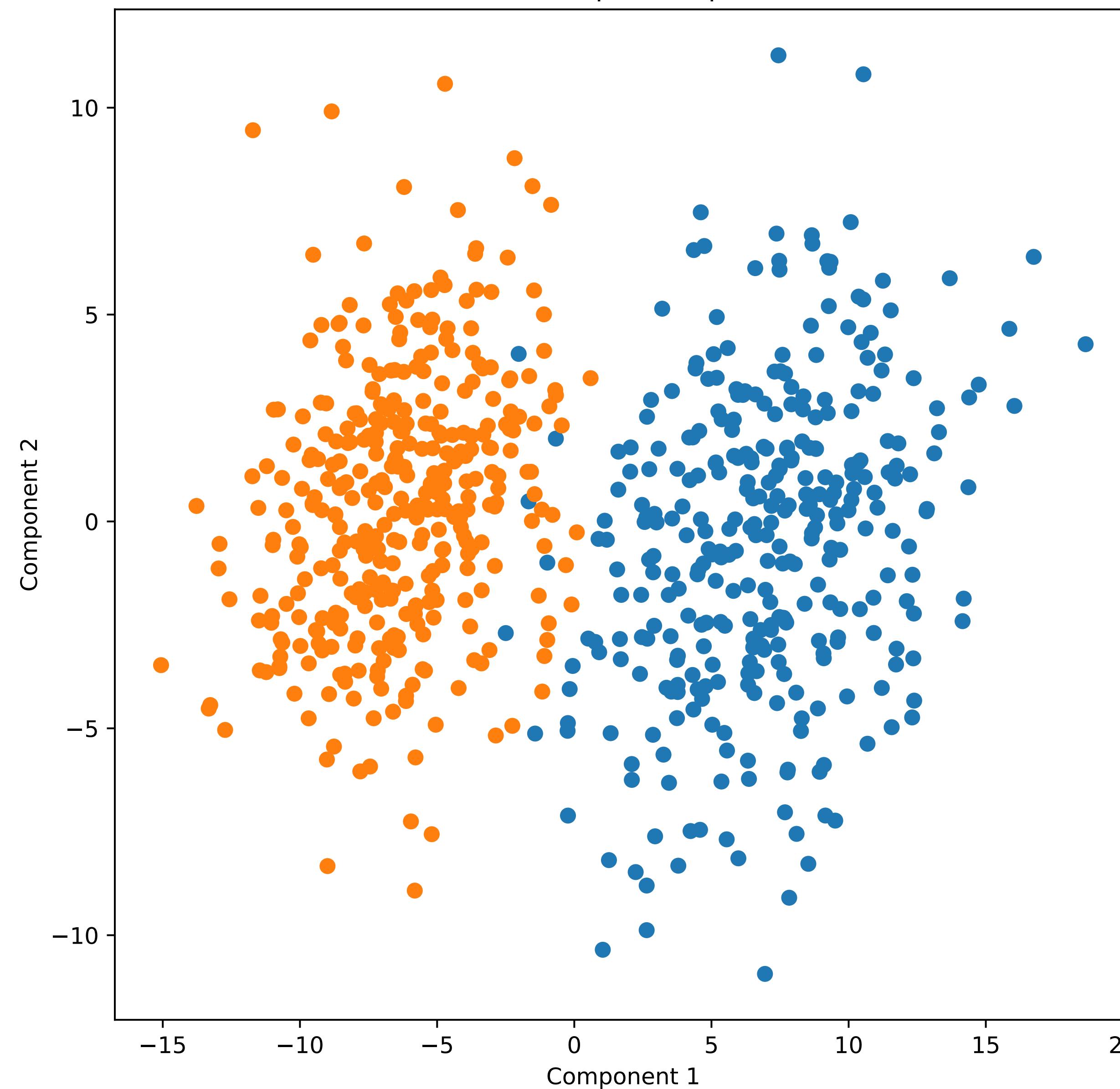
Agglomerative

Divisive





First 2 Principal Components





Questions?

Next: Expectation Maximisation

