Unsupervised Learning

Machine Learning for Data Science 1

Draws inference from data without labeled responses.

Wikipedia: type of ML to find previously undetected patterns in a data with no preexisting labels and with minimal human supervision.

not true

interpretation is central

and so is the human guidance

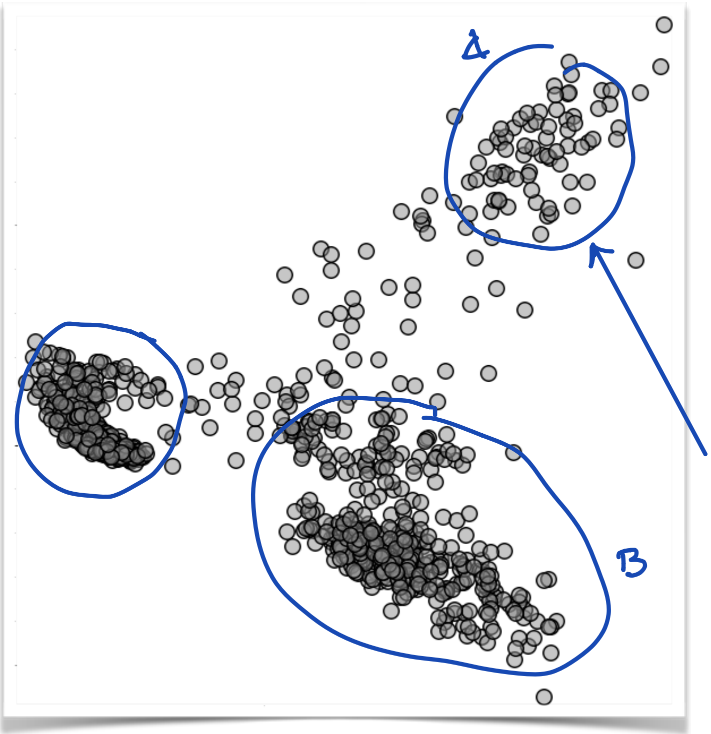
in these approach

choice of parameters

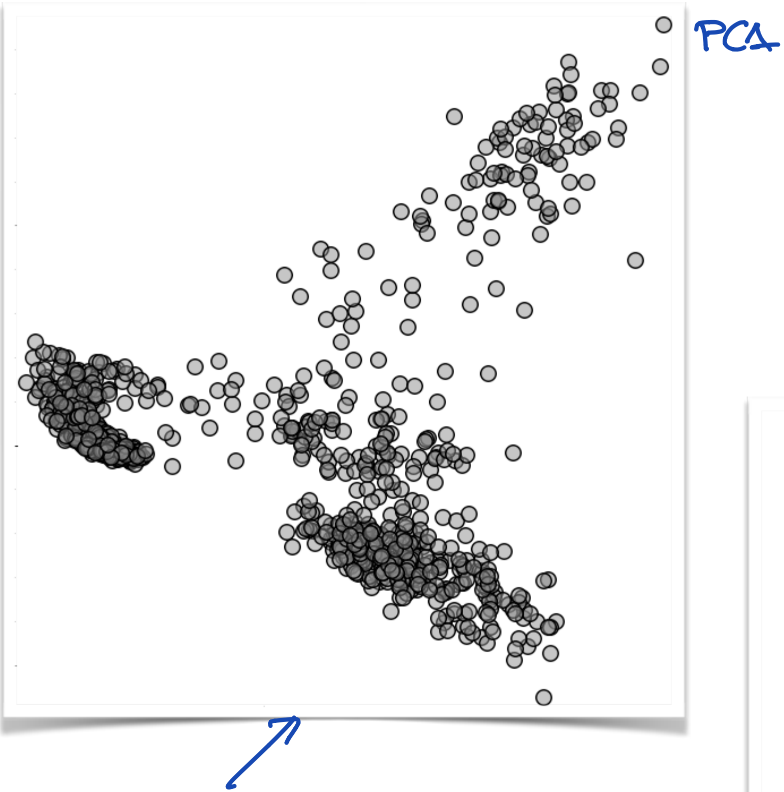
approach

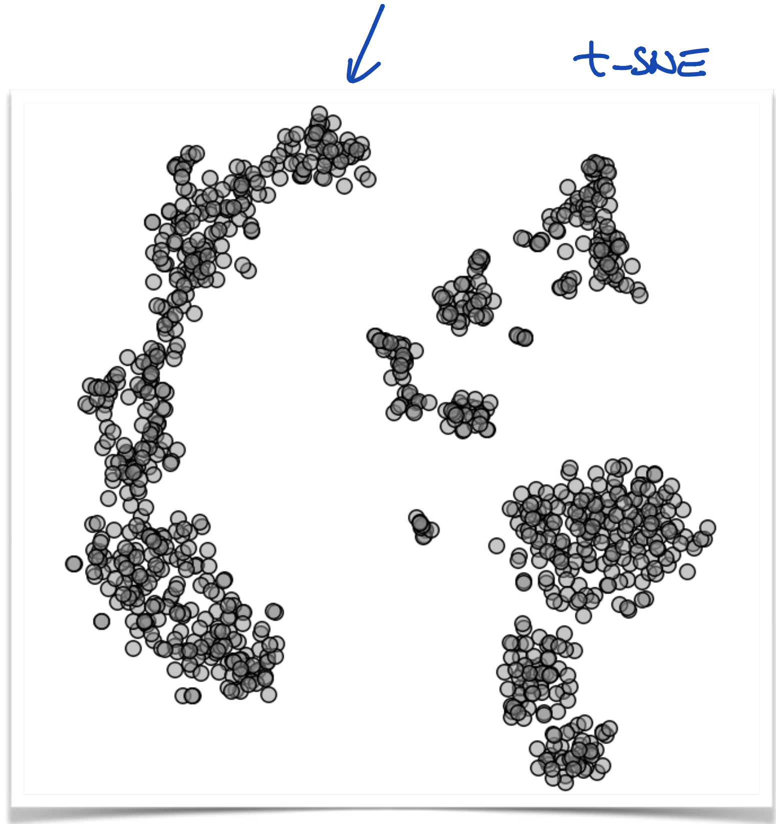
overfitting

overinterpretation



original space of >1000 features
principal component analysis
how many groups do we see
interpretation
is the data that I have enough
knowledge.


PCA


t-SNE

principled approaches to unsupervised learning

dimensionality reduction

- projections (change of coord. sys)
- embedding (new latent space)

2D, 1D

faithfulness

Clustering

grouping of data instances

similarity within

dissimilarity betw. groups

still leaves room for interpretation

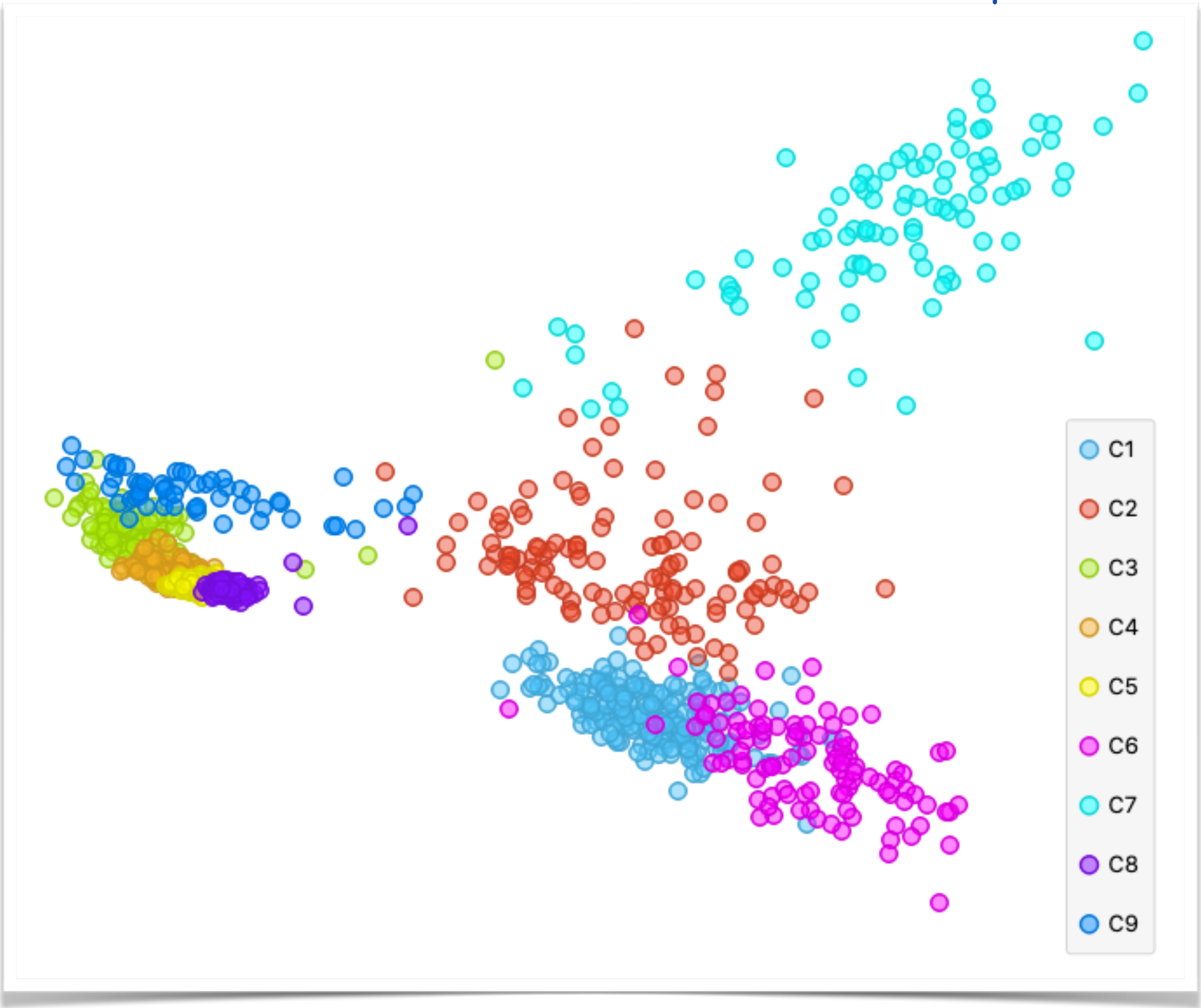
many choices of parameters

in original space

combination

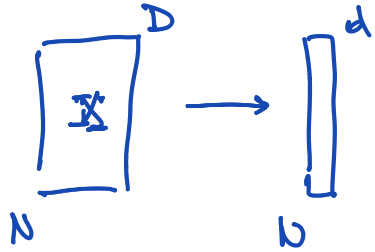
cluster + project, visualize

PCA , >1000



Principal Component Analysis

dimensionality reduction



$D \gg d$

$d = 2 \leftarrow$ visualization

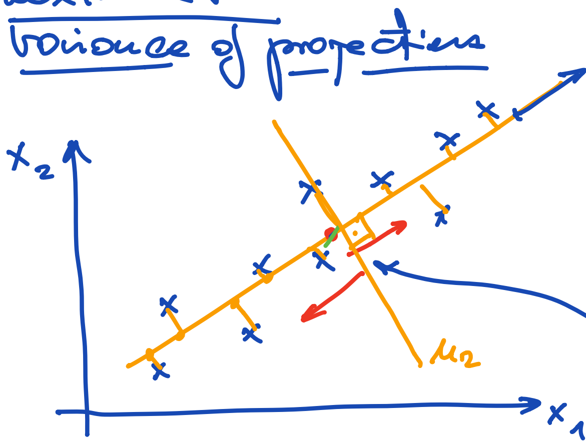
$d = 1 \leftarrow$ sense of time
progression
development

maximize the variance of projections

$u_1 \leftarrow$ direction of projection

$u_1^T u_1 = 1$

a unit vector



$x \in X$

$u_1^T x \in \mathbb{R}$

$u_1^T \bar{x}$, $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$

we are looking for u_1 so that projected data points are maximally dispersed

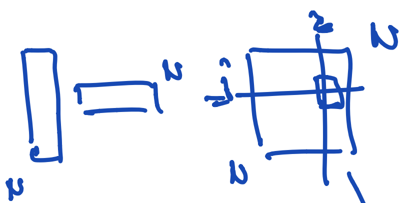
$$\frac{1}{N}\sum_{i=1}^{N}\left(u_1^T x_i - u_1^T \bar{x}\right)^2 = \underline{\underline{Var\left(u_1^T X^T\right)}}$$

$$Var\left(u_1^T X^T\right) = \frac{1}{N}\sum\left(\underline{u_1^T}\, x_i\, \underbrace{u_1^T x_i}_{} - 2\,\underline{u_1^T}\, x_i\, \underbrace{u_1^T \bar{x}}_{} + \underline{u_1^T}\,\bar{x}\, \underbrace{u_1^T \bar{x}}_{}\right)$$

$$\left(u_1^T x_i\right)^T = x_i^T u_1^T$$

$$= u_1^T\left(\frac{1}{N}\sum\left(x_i x_i^T - 2 x_i \bar{x}^T + \bar{x}\bar{x}^T\right)\right)u_1$$

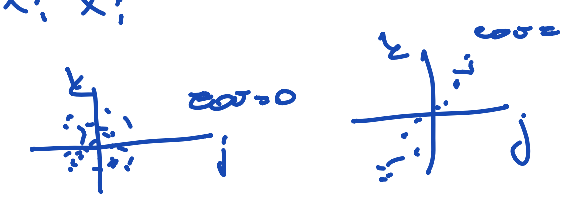
$$= u_1^T\left(\underbrace{\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\bar{x}\right)\left(x_i-\bar{x}\right)^T}_{}\right)u_1$$



$$\sum_{i=1}^{N} x_i^j\, x_i^k$$

K_{xx}

covariance matrix



$$\mathrm{Var}(u_1^T X^T) = \underline{u_1^T \; K \; u_1} \rightarrow \text{For PCA maximize}$$

$$\text{constrain: } \underline{u_1^T u_1 = 1}$$

Lagrange:

$$f(u_1) = \underline{u_1^T K} \, u_1 + \lambda_1 (\underline{1 -} u_1^T u_1)$$

$$\nabla f(u_1) = \underline{K u_1 - \lambda_1 u_1 = 0}$$

$$\underline{K u_1 = \lambda_1 u_1}$$

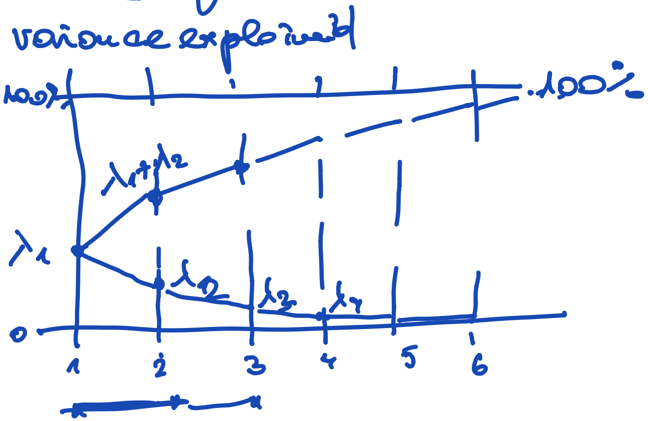
eigenvector

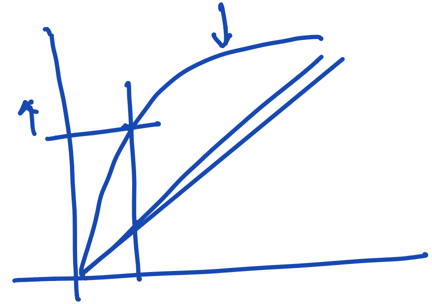
eigenvalue

$$u_1^T K u_1 = \underline{\mathrm{Var}(u_1^T X^T)} = u_1^T \lambda_1 u_1 = \underline{\lambda_1}$$

PCA : 1st component : u_1 of K, λ_1

Scree diagram

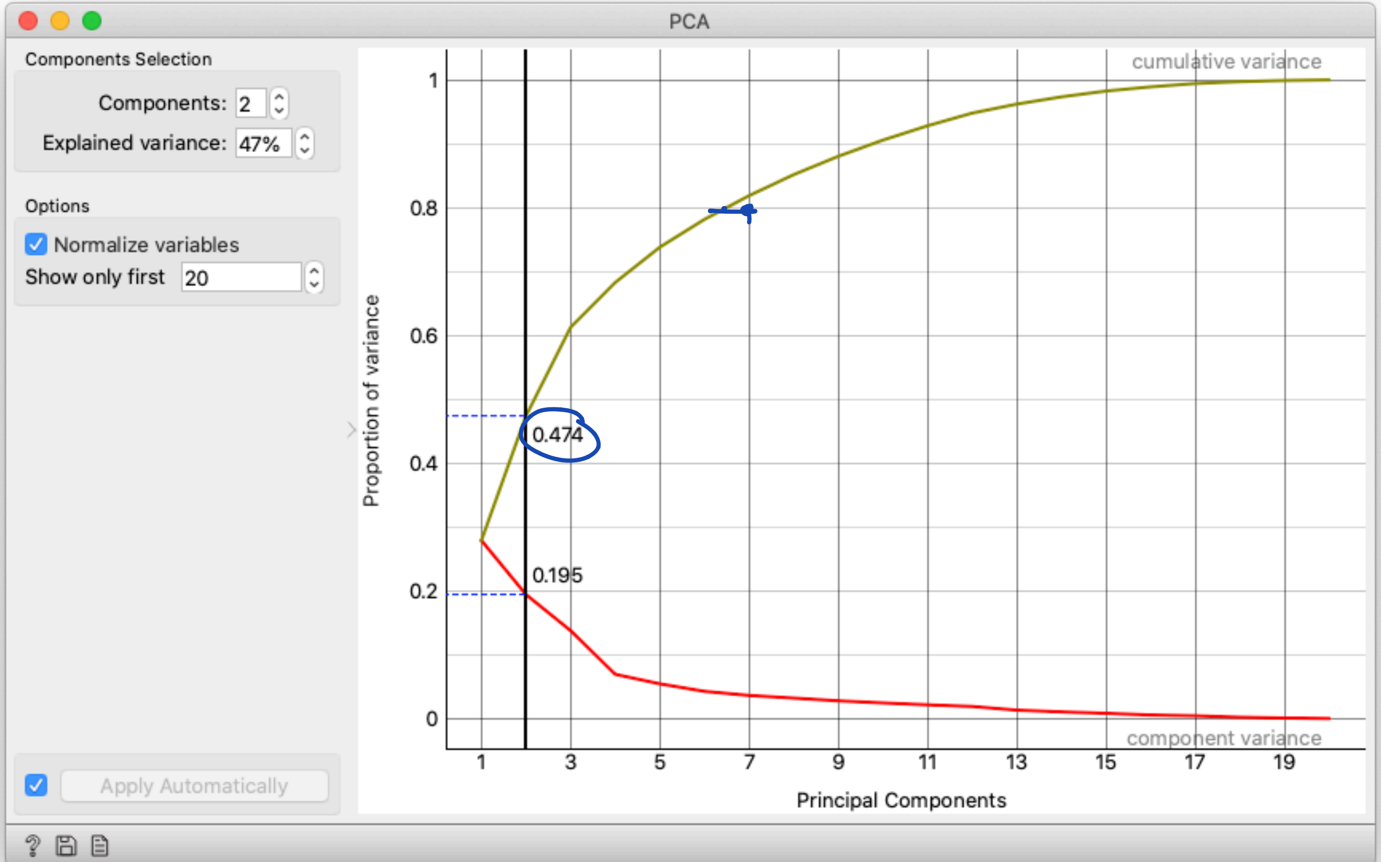


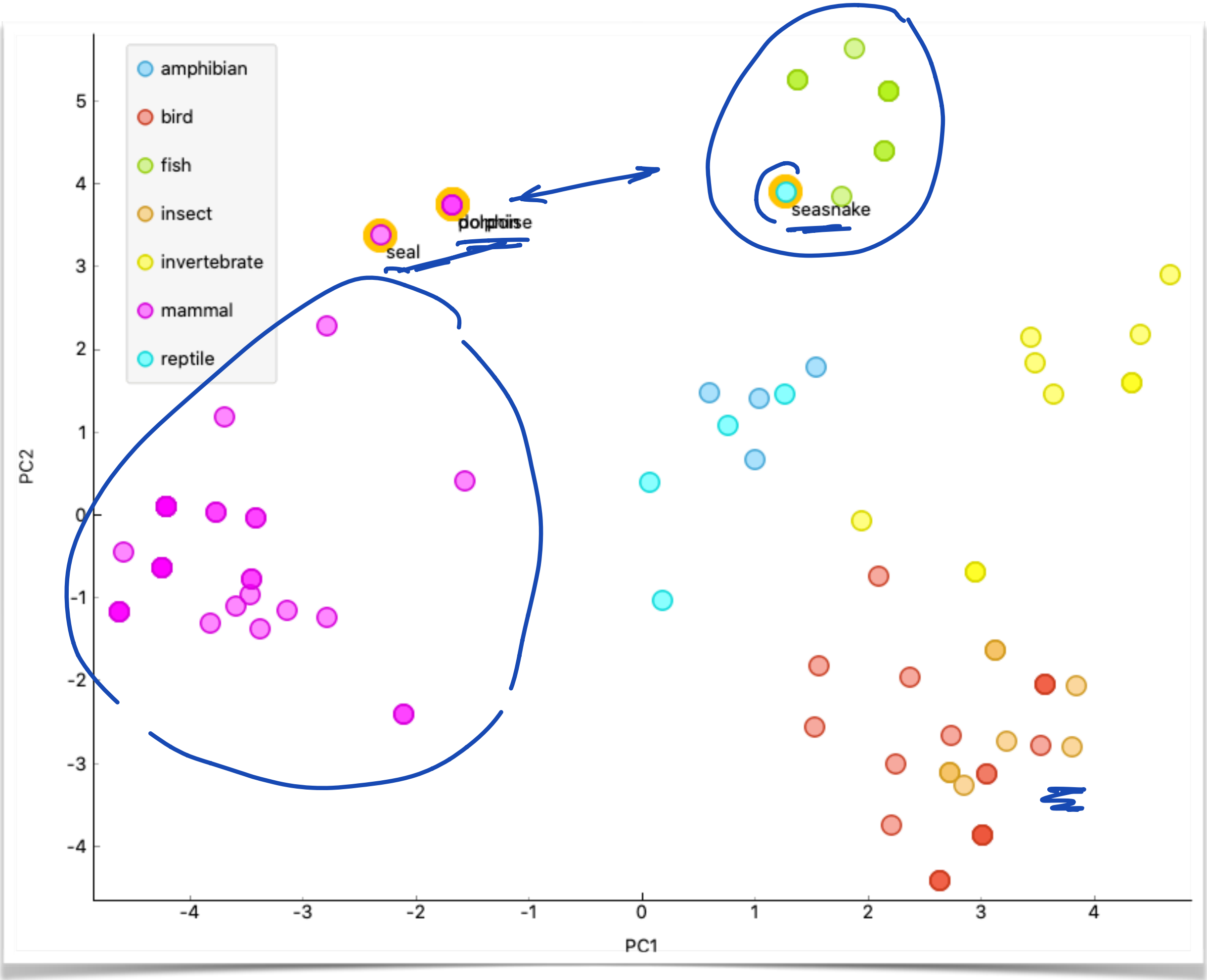


80% , 90%

- how many dimensions?
- are the first 2 dim. "informational"

- power method
- gram-schmidt orthog.


PCA
Components Selection
Components: 2
Explained variance: 47%
Options
Normalize variables
Show only first 20
Apply Automatically
cumulative variance
component variance
Proportion of variance
Principal Components
0.474
0.195
1
0.8
0.6
0.4
0.2
0
1
3
5
7
9
11
13
15
17
19


amphibian
bird
fish
insect
invertebrate
mammal
reptile
seal
dolphin
porpoise
seasnake
PC1
PC2
-4
-3
-2
-1
0
1
2
3
4
5

Singular Value Decomposition

		Drama			SF	
		A star is born	Titanic	When Sally met Harry	Matrix	Predator
women	Mary	1	2	1	0	0
	Eve	5	6	7	0	0
	JoAnn	2	3	1	0	0
	Jim	1	2	1	1	0
men	Fritz	0	0	0	5	7
	Bill	0	0	0	4	2

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$\rightarrow$

1	1	1	0	0
0	0	0	1	1

S truncated decomposition

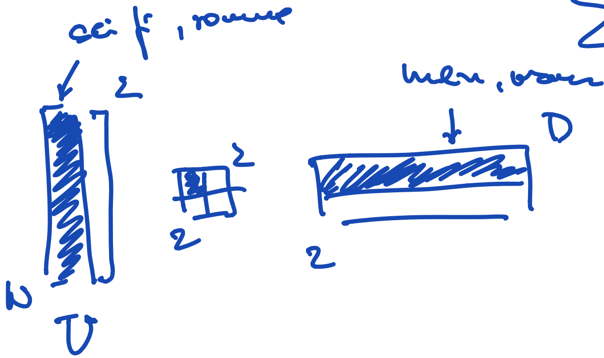
$$\underset{N\times D}{X} = \underset{N\times R}{U}\,\underset{R\times R}{\Sigma}\,\underset{R\times D}{V^T}$$

$$U^T U = I$$

$$V^T V = I$$

Σ : diagonal

$$\hat{X} =$$



$$X = U \Sigma V^T$$

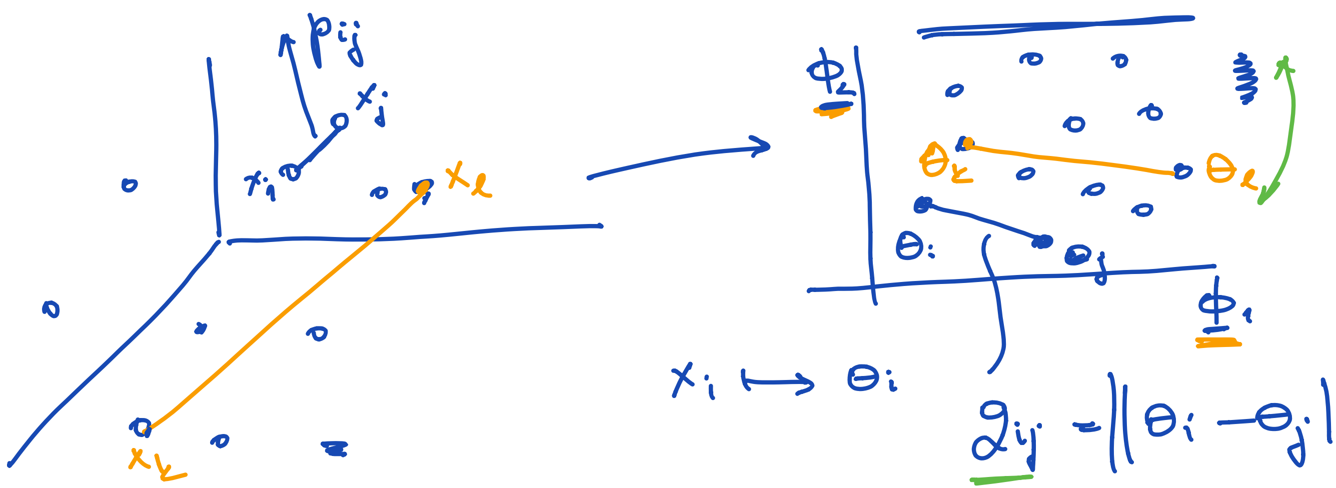
$$X^T = V \Sigma^T U^T = V \Sigma U^T$$

$$X^T X = V \Sigma^T \underbrace{U^T U}_{I} \Sigma V^T = V \Sigma^2 V^T$$

$$X^T X V = V \Sigma^2 V^T V = V \Sigma^2$$

PCA : projection

embedding : Multidimensional Scaling MDS



$x_i \longmapsto \theta_i$

$$q_{ij} = \| \theta_i - \theta_j \|$$

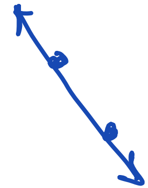
$$p_{ij} = \| x_i - x_j \|$$

any kind of distance

Preserve the distances !

$$J(\Theta) = \sum_{i \neq j}^{N} \left(p_{ij} - q_{ij} \right)^2 \quad : \text{ minimize}$$

$$\left[\; X \;\right]^{D}_{N=100} \rightarrow \left[\; \Theta \;\right]^{2}_{200}$$



$\swarrow$ 200

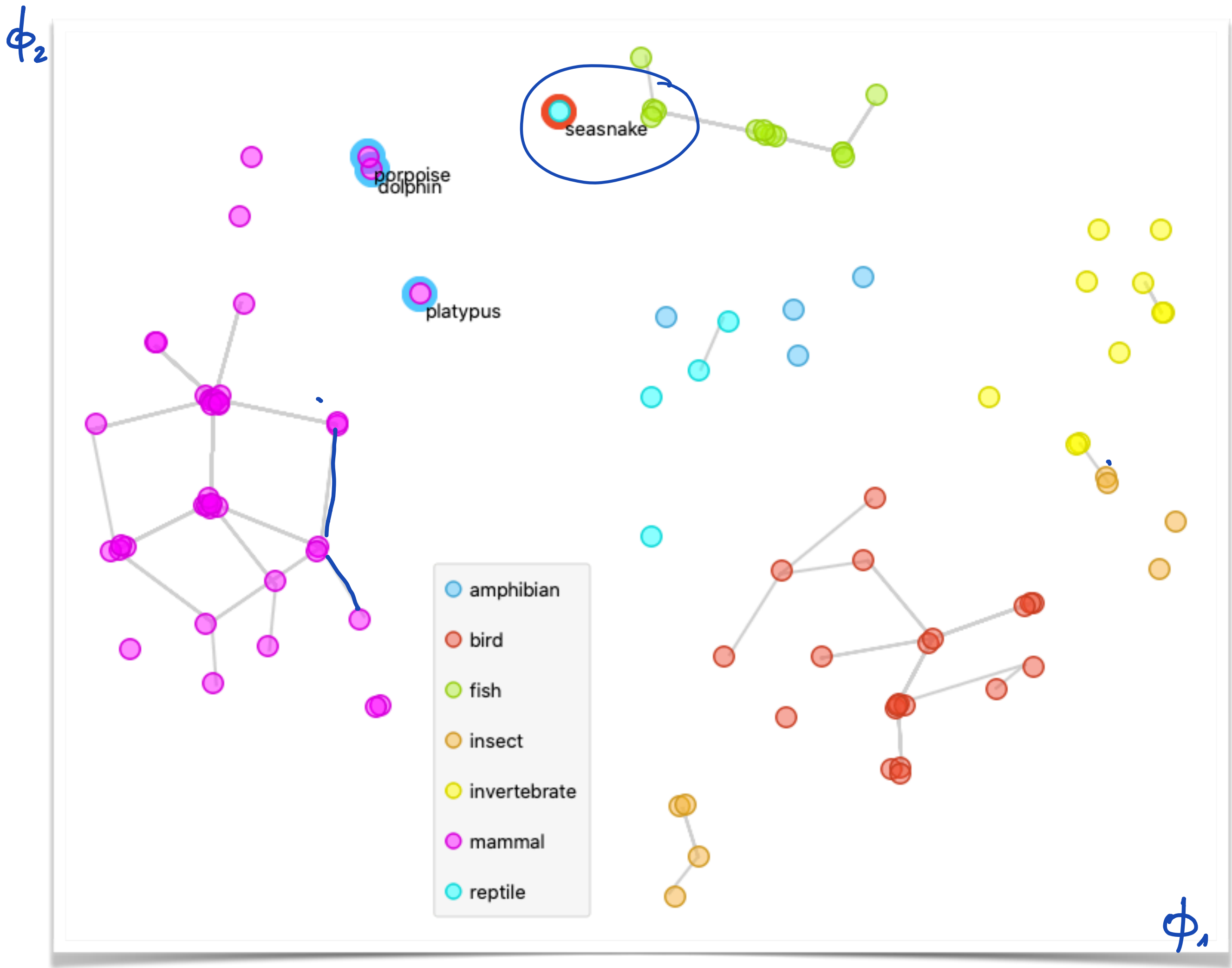
$$\frac{\partial J(\theta)}{\partial \phi_1^{(i)}} = \sum_{i \neq j} \underbrace{\frac{\partial J(\theta)}{\partial g_{ij}}}\;\underbrace{\frac{\partial g_{ij}}{\partial \phi_1^{(i)}}}$$

$$= 2 \sum_{i \neq j} \underbrace{\left(g_{ij} - p_{ij} \right)}_{\text{error}} \; \underbrace{\frac{\phi_1^{(i)} - \phi_2^{(j)}}{g_{ij}}}$$

| direction of change

Solution: majorization, SMACOF

$\hookrightarrow$ speedup convergence



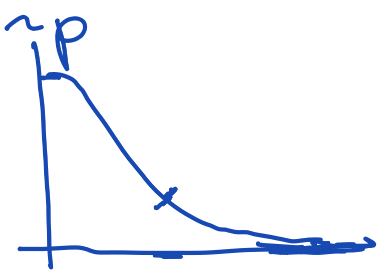
ϕ_2
seasnake
porpoise
dolphin
platypus
amphibian
bird
fish
insect
invertebrate
mammal
reptile
ϕ_1

Stochastic Neighbor Embedding - SNE

2002 Hinton & Roweis

2008 van der Maaten t-SNE

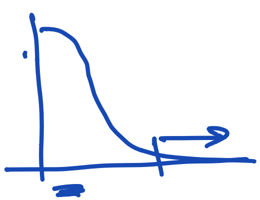
Idea: datapoints close to each other in the original space should be close in embedding



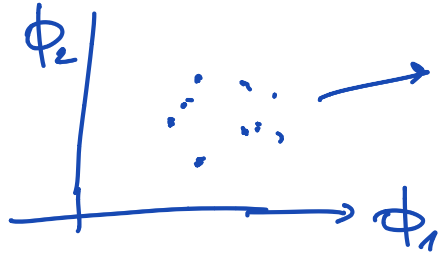
$$p_{j|i} = \frac{e^{-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}}}{\sum_{k \neq i} e^{-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}}}$$

perplexity

original space



to latent space



$$q_{j|i} = \frac{e^{-\|\phi_i - \phi_j\|^2}}{\sum_{k \neq i} e^{-\|\phi_i - \phi_k\|^2}}$$

SNE

$$J(\theta) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} = \sum_i KL(P_i \| Q_i)$$

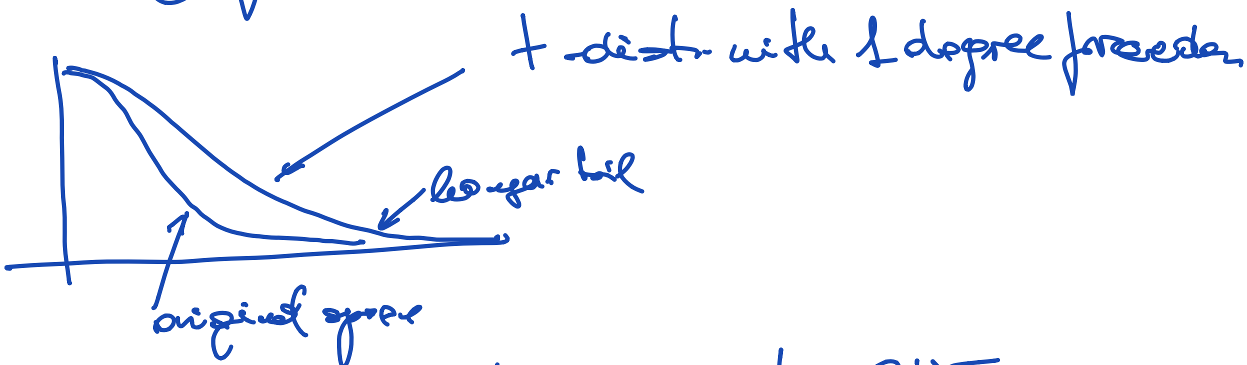
$$\frac{\partial J}{\partial \theta^{(i)}} = 2 \sum_j \left(p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j}\right)\left(\theta^{(i)} - \theta^{(j)}\right)$$

SNE

multi dimension space → two dimensional space

1000 features → 2

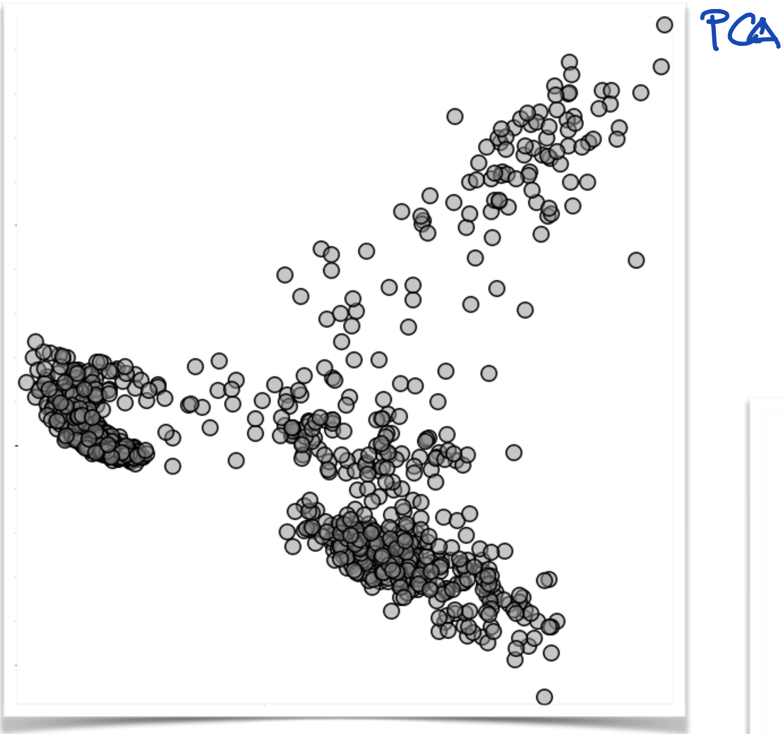
The crowding problem

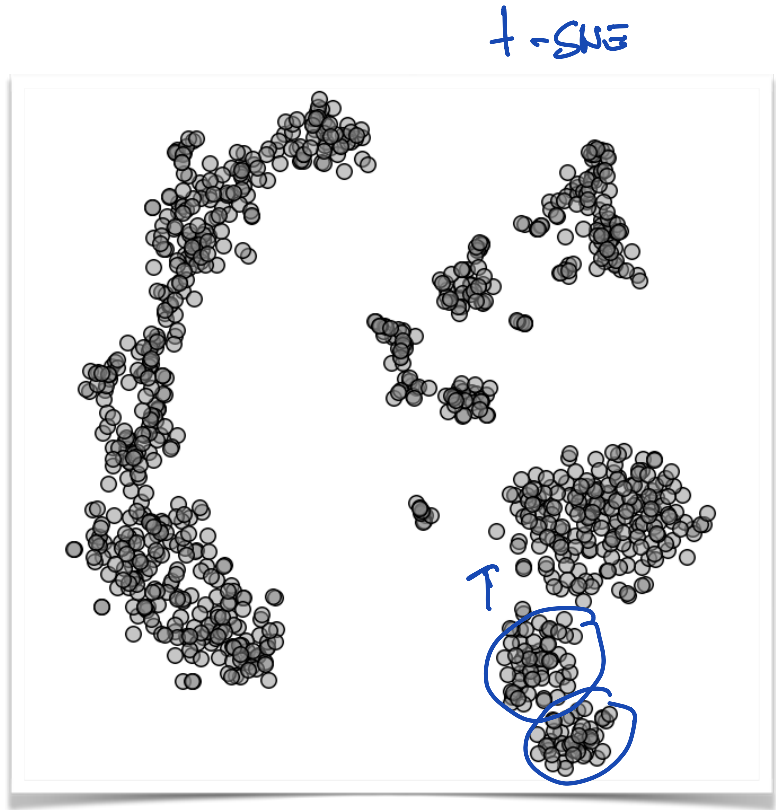


$$q_{ij} = \frac{\left(1+\|x_i - x_j\|^2\right)^{-1}}{\sum_{k \neq i}\left(1+\|x_i - x_k\|^2\right)^{-1}}$$

t-SNE

$$\frac{\partial J}{\partial \theta^{(k)}} = 4 \sum \underbrace{\left(p_{ij} - q_{ij} \right)}_{\text{error}} \underbrace{\left(\phi_i - \phi_j \right)}_{\text{direction}} \underbrace{\left(1 + \| \phi_i - \phi_j \|^2 \right)^{-1}}_{\text{weight}}$$

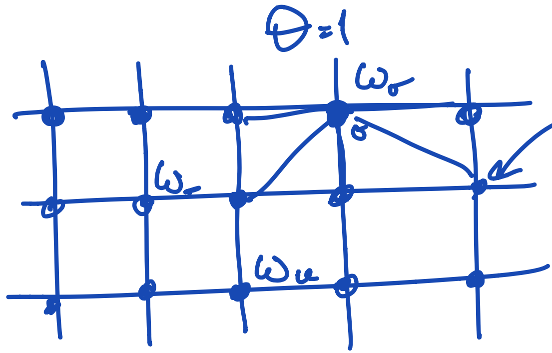

PCA


t-SNE
T

Self-Growing Maps

Kohonen maps

networks



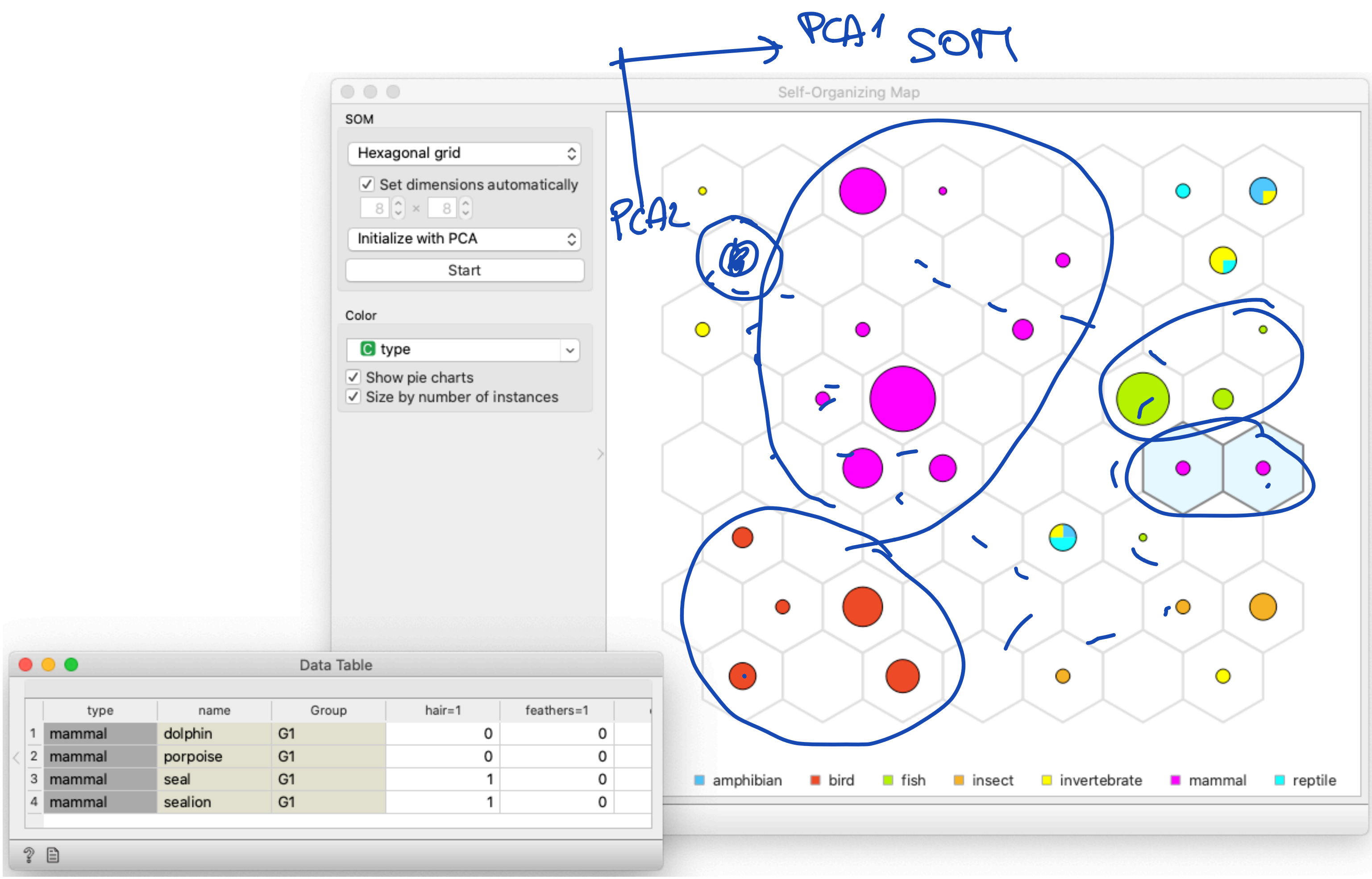
$x \in \underline{X}$

w_v : prototype

1. initialize w_v
2. randomly pick $x \in X$
3. find noble u.

$$u = \arg\max_{v} \|x - w_v\|$$, we place x in node u

4. $$w_v \leftarrow w_v + \Theta(\underline{u}, \underline{v}) \cdot \alpha \, (\underline{w_v} - \underline{x})$$
5. if stopping condition not met
   goto 2


PCA1
SOM
PCA2
Self-Organizing Map
SOM
Hexagonal grid
Set dimensions automatically
8
8
Initialize with PCA
Start
Color
type
Show pie charts
Size by number of instances
amphibian
bird
fish
insect
invertebrate
mammal
reptile
Data Table
type
name
Group
hair=1
feathers=1
1 mammal dolphin G1 0 0
2 mammal porpoise G1 0 0
3 mammal seal G1 1 0
4 mammal sealion G1 1 0